

RESEARCH

Open Access

# Automated linear regression tools improve RSSI WSN localization in multipath indoor environment

Frank Vanheel<sup>1\*</sup>, Jo Verhaevert<sup>1</sup>, Eric Laermans<sup>2</sup>, Ingrid Moerman<sup>2</sup> and Piet Demeester<sup>2</sup>

## Abstract

Received signal strength indication (RSSI)-based localization is emerging in wireless sensor networks (WSNs). Localization algorithms need to include the physical and hardware limitations of RSSI measurements in order to give more accurate results in dynamic real-life indoor environments. In this study, we use the Interdisciplinary Institute for Broadband Technology real-life test bed and present an automated method to optimize and calibrate the experimental data before offering them to a positioning engine. In a preprocessing localization step, we introduce a new method to provide bounds for the range, thereby further improving the accuracy of our simple and fast 2D localization algorithm based on corrected distance circles. A maximum likelihood algorithm with a mean square error cost function has a higher position error median than our algorithm. Our experiments further show that the complete proposed algorithm eliminates outliers and avoids any manual calibration procedure.

**Keywords:** algorithm design and analysis, correlation and regression analysis, wireless sensor networks, localization

## 1 Introduction

Wireless sensor networks (WSNs) are expected to offer a cheap solution not only for monitoring and control applications, but also for more advanced applications like location-based services (e.g., tracking and tracing of persons and objects, indoor guiding of persons in complex buildings, offering location-based information, etc.). The main requirement for mass deployment of WSNs and corresponding services is the easy installation and configuration, which is realized by the introduction of self-organizing and auto-configuration mechanisms. For location-based services, the same requirements are valid: a cheap technology in combination with a simple deployment strategy avoiding complex and time-consuming manual configuration and calibration procedures. This article presents a new approach for developing a localization algorithm, and has validated the algorithm in a real-life test bed. The approach is based on the automatic selection of anchor nodes for

received signal strength indication (RSSI)-based indoor localization, and hence avoids any manual calibration.

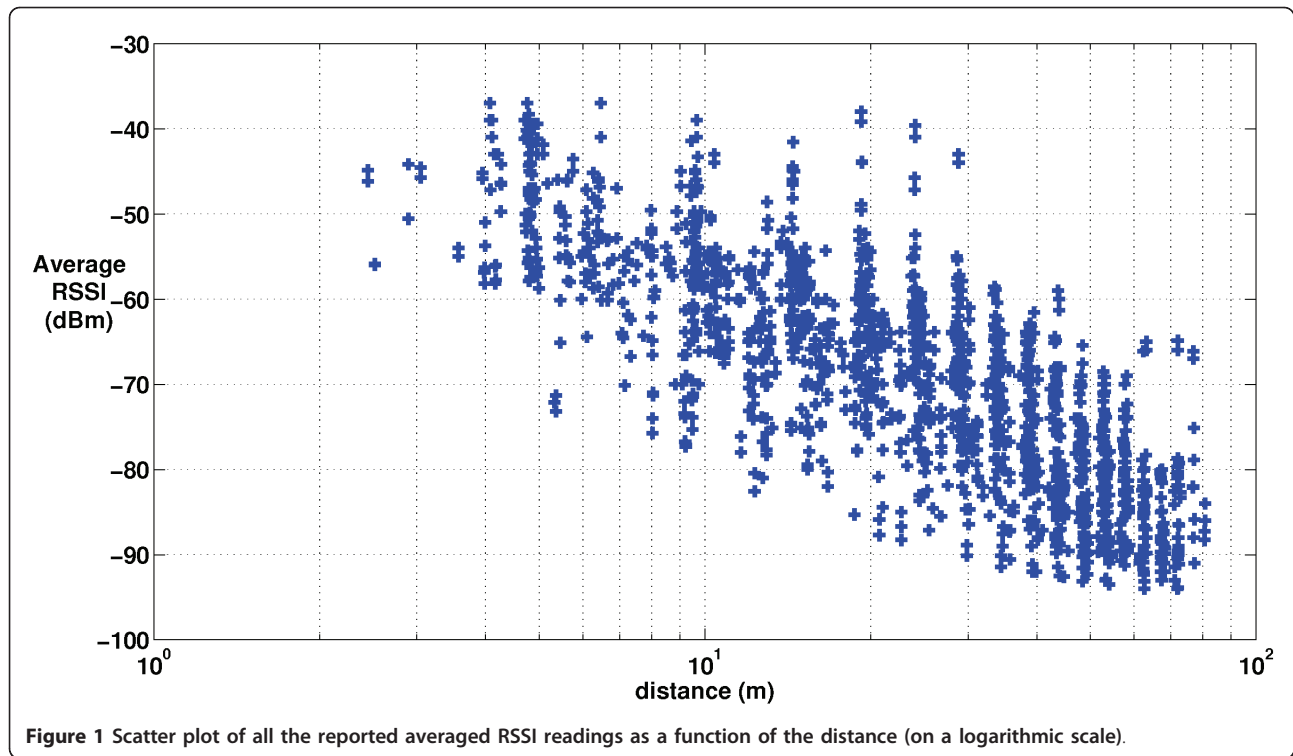
RSSI-based localization is two-sided [1]. The first method explores the physical (theoretical) relationship between the RSSI and the distance. Free space models and two ray ground models are the key words. The second (experimental) method uses fingerprinting: a RSSI database is filled with measurement records during an extensive calibration phase, and the location is estimated by fitting the measured RSSI to this database. An example of a mature application of this kind is Mote-Track [2]. In the presence of severe multipath fading with multiple reflections (more than 30 are reported in the literature [3]), the relationship between RSSI and distance is extremely hard to model, making both methods ineffective. Some authors conclude that the large amount of characterization will make the use of signal strength approaches with low power radios practically impossible [4].

An initial look at Figure 1, where RSSI versus the distance (expressed in m) between all sending and receiving nodes of our building is plotted on a semi-logarithmic scale, confirms the dominance of multipath fading in indoor environments. Basically, the graph

\* Correspondence: frank.vanheel@ugent.be

<sup>1</sup>Faculty of Applied Engineering Sciences, University College Ghent, Ghent, Belgium

Full list of author information is available at the end of the article



consists of 1942 RSSI-log(distance) pairs, where the distance is expressed in meter. 47 nodes broadcast 240 packets to all other nodes, and the average of the RSSI reported by the receiver and its distance to the sender gives one point. Packets below the sensitivity level of a receiver are not reported and thus not presented in the graph. The configuration itself is explained in Section 4. For example, at a distance of 24 m, RSSI values between -40 and -84 dBm are encountered. Alternatively, an RSSI of -65 dBm corresponds to distances ranging between 5.5 and 77 m which actually covers almost the entire building. This large RSSI variability is also found in other experimental studies in industrial indoor environments [5]. It is obvious that, in such realistic environment, physical relationships cannot be applied as such. We therefore use standard statistical tools to solve this problem. We assume a pre-existing sensor network with a large number of nodes, which is a realistic scenario for future dynamic wireless indoor environments. In a first step we select well-behaving 'anchor nodes' [6] from all the active nodes and calibrate them to their individual propagation parameters according to their underlying physical behavior. To prove the concept of our method, we use the unselected nodes (non-anchors) as localization targets. However, our technique can also be used for the localization of other targets, for which no location information is available. Our automated preprocessing step

polishes the data and rejects too small and too large distance circles and uses a fast maximum likelihood algorithm on the distance. Finally, we propose a fast 2D localization algorithm based on the corrected distance circles and compare the results with the more conventional maximum likelihood on the position algorithm with a mean square error cost function.

This article is organized as follows: In Section 2, an overview of related work is given. In Section 3, the realistic wireless environment is described. Section 4 presents the discovery of multipath fading in a corridor and discusses the impact on power levels. In Section 5, preprocessing of the measurements is treated, followed by the positioning in Section 6. Section 7 compares the test results of the proposed algorithm with those of a more conventional maximum likelihood algorithm. Finally, in Section 8, conclusions are presented.

## 2 Related work

Possible range-based localization methods are angle of arrival [7], time of arrival [8], time difference of arrival [9], and RSSI. In this article, we only consider RSSI, because this can be implemented with inexpensive hardware. This means that the sensors calculate a distance from the physical RSSI measurements. Range-free localization methods do not use physical parameters but work with the content of a message and are also not treated here.

There are different technologies used for indoor localization: ultrasonic [10], infrared [11], Bluetooth [12], RFID [11,12], WiFi [13,14], UWB [15-17], and a combination of technologies [18,19]. These technologies all have not only their respective advantages, but also shortcomings. Ultrasonic and infrared waves cannot propagate through walls. Bluetooth has not only a limited range, but also a restricted number of devices. Power consumptions of both RFID and WiFi are large compared with IEEE802.15.4 sensors. Furthermore, RSSI measurements on WiFi are tricky, since the absolute accuracy on the RSSI is not specified in the standard [20], making the RSSI manufacturer dependent. For IEEE802.15.4, however, there are very few chipmakers. Most devices use the CC2420 chip from Chipcon with a dynamic range of 100 dB, and the datasheet specifies the RSSI linearity (the relationship between RSSI and received input power) within the limits of  $\pm 3$  dB.

UWB is still a new technology following the IEEE802.15.4a standard. Only few devices are available [21] (at reasonable cost), and only very recently have theoretical performance bounds of wideband localization been described [22-24]. UWB empirical results on localization in harsh multipath environment are also brand new [25,26].

Distance tolerances of these wireless indoor positioning techniques can be found in [27]. To our knowledge, an overview of empirical localization accuracies based on IEEE802.15.4 devices does not exist yet, but these studies are proliferating and some individual results can be found in [28,29].

Tremendous effort is already put in localization algorithms [30-33] in general. RSSI-based localization can be found with the Bayesian estimators [34-36] and maximum likelihood estimation [37]. These theoretical estimators have in common the ability to estimate the propagation parameters in their localization algorithm itself. We will not follow this approach and only feed good empirical input to our simulator. Preprocessing the data according to the known physical behavior will result in better models at the input and hence at the output of the positioning algorithm. The best anchor nodes are carefully selected with automated standard linear regression tools. Now the localization algorithm can be optimized on spatial constraints only.

The techniques described in [1,38,39] show some similarity with ours, but we push the limits one step further by an automated selection of anchors in a larger sensor network based on standard linear regression techniques.

The authors of [1] also suggest a new approach for range-based localization using RSSI. They also do not use fingerprinting and avoid the expensive training phase of MoteTrack-like algorithms and estimate the propagation parameters using the RSSI measurements.

Our major improvement is that we use more than three anchors for the calibration and training phase at no additional time cost.

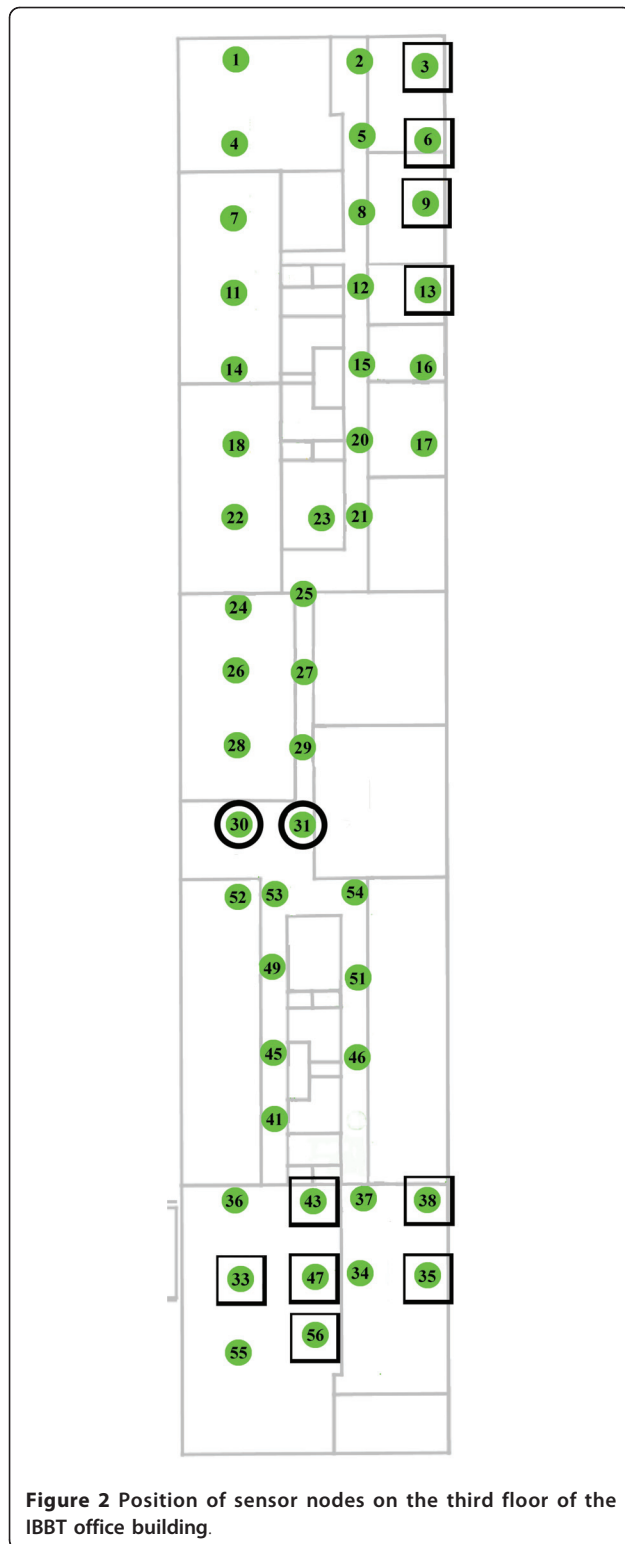
With the authors of [38], we share the experimental analysis of RSSI-based localization. Also, the linear fitting is common. Our experiments, however, are performed using IEEE802.15.4 devices and use a larger dataset in a larger realistic environment.

The calibration method described in [39] is the starting point of our study. In our article, a selection method of anchors is added and a localization algorithm is presented.

Our work completely exploits the observed correlation between the quality of anchors and the linearity of the calibrated path loss model as described in [34] and adds a new, fast, and simple algorithm using the basics not only of the underlying physics (propagation model and multipath fading), but also of linear regression (standard error, confidence, and expectation intervals).

### 3 Used equipment

The Interdisciplinary Institute for Broadband Technology (IBBT) iLab.t Wireless Lab or W-ilab.t is an extensive wireless mesh and sensor network infrastructure installed at IBBT office premises in Ghent (Belgium), including meeting rooms, classrooms, offices, and corridors. The W-ilab.t test bed has more than 200 Tmote Sky nodes equipped with a Chipcon CC2420 radio chip operating in the 2.4-GHz frequency band. The nodes are spread over three floors on a grid measuring  $12 \times 80 \text{ m}^2$ . The dimensions of the third floor are  $16.8 \times 90.0 \text{ m}^2$  or  $1512 \text{ m}^2$ . On the third floor, there are 56 nodes. All the nodes are mounted at the same height beneath a highly conducting ceiling. The central internal walls are constructed of plywood, while walls in the vicinity of the staircases and elevator are made of concrete. At the elevator, a lot of metal has been used. One side of the longest corridors is concrete and the other plywood. Only the third floor is discussed in this article. In Figure 2, this floor is shown with the positions of the 47 active nodes and its four narrow but long corridors. The anchors with numbers 10, 19, 32, 39, 40, 42, 44, 48, and 50 were closed down for maintenance during our experiment. Every node is connected to a corresponding environment emulator (EE). This component can measure real-time energy consumption, emulate battery depletion, and generate artificial sensor data, hence emulating real-world application scenarios. The EE is connected to an intermediate node or iNode. The iNodes are further connected to a central management server via ethernet and are responsible for the configuration and control of the sensor nodes. The iNodes are also connected with two IEEE802.11 (a/b/g) radios, and allow WiFi testing. In this article, we will only focus on



the sensor nodes. Power is fed across ethernet through power over ethernet network switches. The control software supports both Tiny OS 1 and Tiny OS 2. The iNodes can be synchronized with a time server. A time

accuracy of 2  $\mu$ s is obtained and used for data logging. More details about our test bed can be found in [40].

On the iNode, we have a software tool called RadioPerf which is, in terms of functionalities, very similar to the iPerf tool. The tool includes a packet generator and Java-based configuration and visualization of measurements reported to the central data server. For this study, the following items are collected in a log file: the number of packets sent, the number of packets sent with an error, the minimum estimated noise floor, the average estimated noise floor, the maximum estimated noise floor, an identifier of the node generating the packets, the number of packets received, the number of packets lost, the minimum link-quality indication (LQI), the average LQI, the maximum LQI, the minimum RSSI, the average RSSI, the maximum RSSI, a time stamp, the node reporting these values, and a report sequence number. Our dataset is available at <http://users.ugent.be/~inmoerma/LiReFloA>.

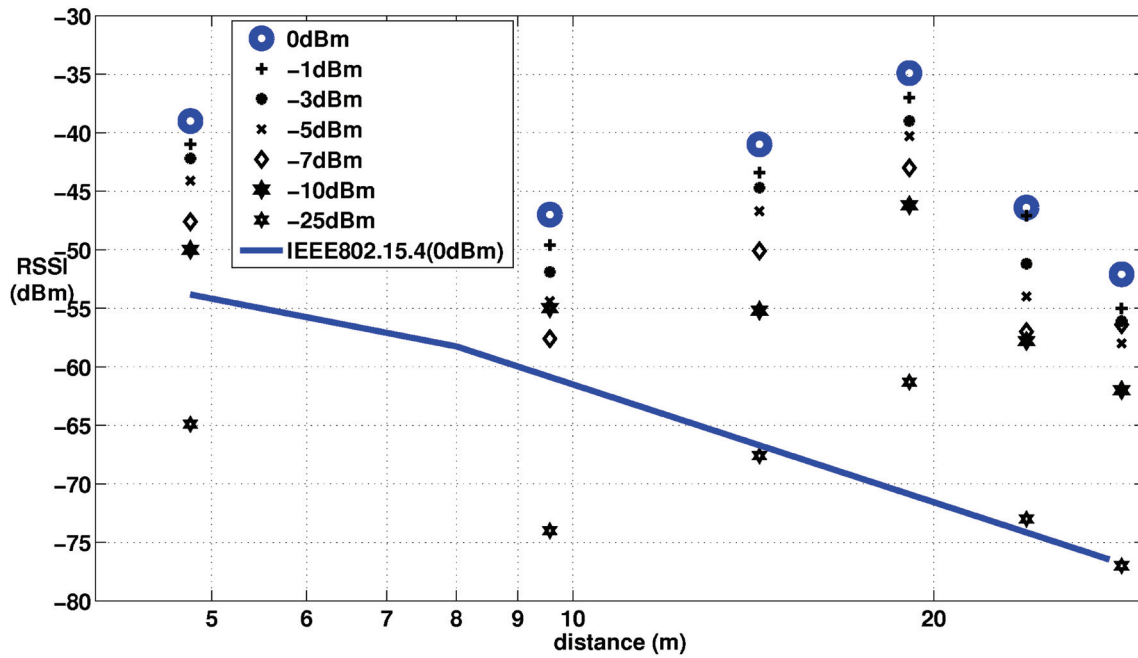
#### 4 Multipath fading

Every node broadcasts 240 packets. Every packet consists of 100 bytes, and the inter packet delay is 25 ms. Transmission is at channel 26 to avoid WiFi interference [41]. Upon swapping sending nodes, the test bed remains idle for 8 ms. The test is repeated for transmit power levels of 0, -1, -3, -5, -7, -10, and -25 dBm. Every second, available data were sent to our log file. Figure 1 was constructed with the average RSSI values of packets sent at full power (0 dBm). The distances between the nodes are known.

In order to explain the big differences in Figure 1, the longest corridor is selected, consisting of nodes 2, 5, 8, 12, 15, 20, and 21. These nodes are collinear, and a clear line of sight (LOS) exists.

Figure 3 shows the RSSI in function of the distance on a semi-logarithmic scale between node 2 and the others at different transmit power levels. The graph also contains the RSSI predicted by the model for a sender emitting at 0 dBm according to (1) and (2) (found in the IEEE802.15.4 standard and described in Section 5.1). The measured RSSI does not decrease monotonously with distance: e.g. for a transmit power of 0 dBm around a distance of 20 m, a RSSI of -35 dBm is measured and at the shorter distance of about 10 m, a RSSI of -47 dBm is observed. Furthermore, at the distance of 20 m, the model in the IEEE802.15.4 expects a RSSI of -70 dBm, giving a difference of 35 dB with the measurement. The measured signal is much too strong, even at small distances. This effect occurs in all corridors and cannot be explained by many theories:

1. First walls have attenuation and what we notice is a stronger signal.



**Figure 3** Capricious LOS RSSI multipath fading in corridors for node 2 at different power levels.

2. Second the standard deviation on measurements due to time-fading is only a few dB [42], and what we see is a very large spread.
3. Third the RSSI accuracy in the CC2420 datasheet is much better than the measured spread.
4. Fourth fluctuations in transmit power are not an issue because Figure 3 only considers one transmitter (node2).
5. Finally, the spread on receiver sensitivity of the CC2420 also does not exceed a few dB, and is therefore less than the observed spread.

The waveform with minima and maxima is also found in other practical experiments in corridors [43] and can only be justified with diffraction and the many constructive scatterings and reflections by the conductive metal ceiling and nearby walls in these corridors, acting as “street canyons” [44].

## 5. Preprocessing

Table 1 gives a schematic description of the preprocessing algorithm. Each procedure will be explained in the following subsections.

### 5.1 Selection of anchors based on the linearity of their calibrated path loss model

Figure 1 looks like a wide monotonously decreasing stripe clipped by a horizontal line at the bottom because of the sensitivity of the receiver. This suggests there still

remains a linear relationship between RSSI and the logarithmic distance.

The free space model suggests a path loss proportional to the square of the distance. The IEEE802.15.4 standard [45] recalculates this path loss to (1) and (2):

$$pl(d) = 40.2 + 20\log_{10}(d) \quad \text{for } d < 8 \text{ m} \quad (1)$$

$$pl(d) = 58.5 + 33\log_{10}(d/8) \quad \text{for } d > 8 \text{ m} \quad (2)$$

with  $pl(d)$  being the path loss in dB at a distance  $d$  (in m) from the receiver. Considering that the path loss is the (logarithmic) difference between sending power and receiving power and emitting at 0 dBm and reading the received power (RSSI) in dBm, we instantaneously get for (1), an intercept point of -40.2 and a slope of -20. For distances greater than 8 m, the standard proposes a steeper slope (-33) and thus proposes a model based on the two ray ground model. It is noted that this is invalid for frequencies other than 2.4 GHz. It is noted that, for WiFi, the RSSI is not calibrated in dBm [20], and an extra step is needed to include the relationship between RSSI and dBm.

In this article, we will calibrate the test bed by determining the intercept point and slope with the linear regression technique, as described in [39].

Based on the RSSI and the (known) distances, we calculate intercept point, slope factor, and the  $r$ -squared (RSQ) values for each sender (and all other nodes

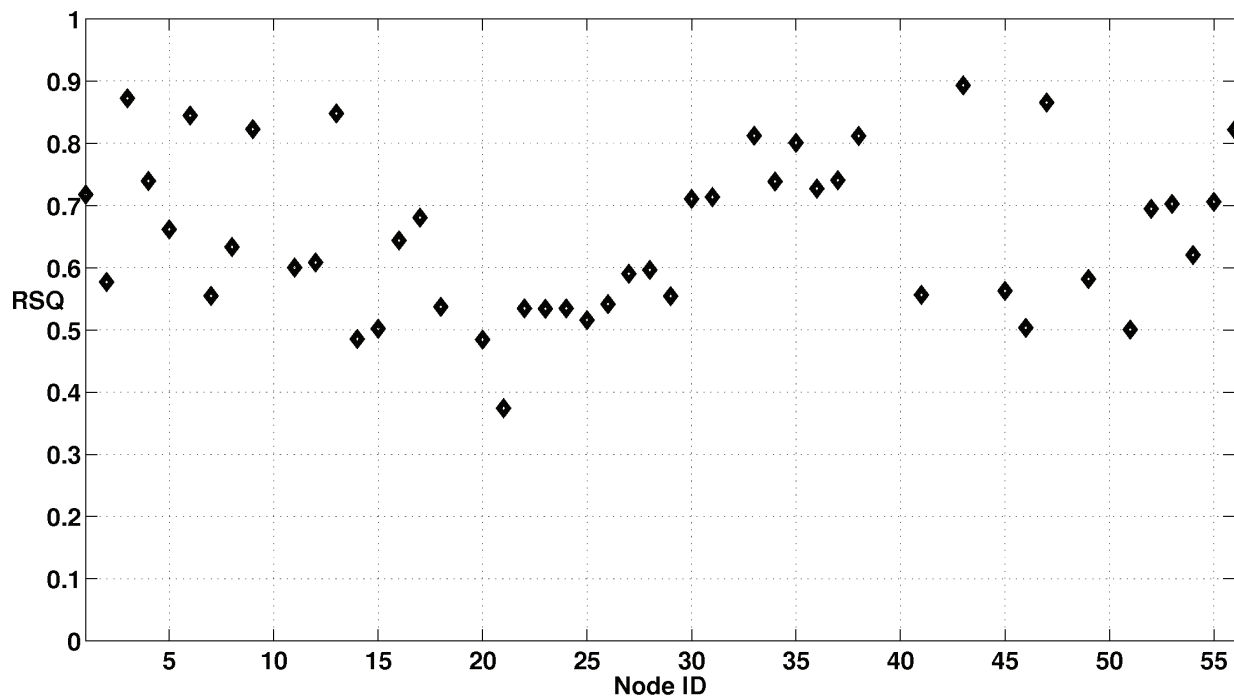


**Table 1 Schematic description of the preprocessing**

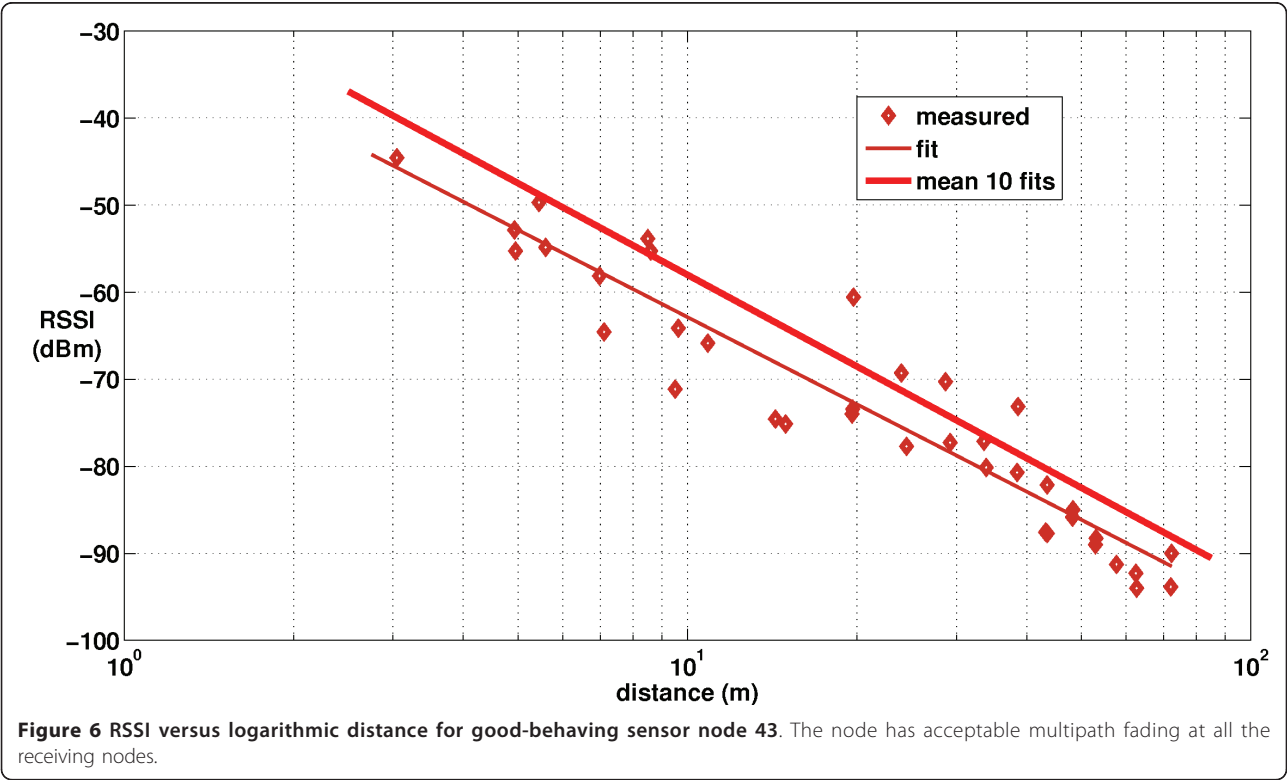
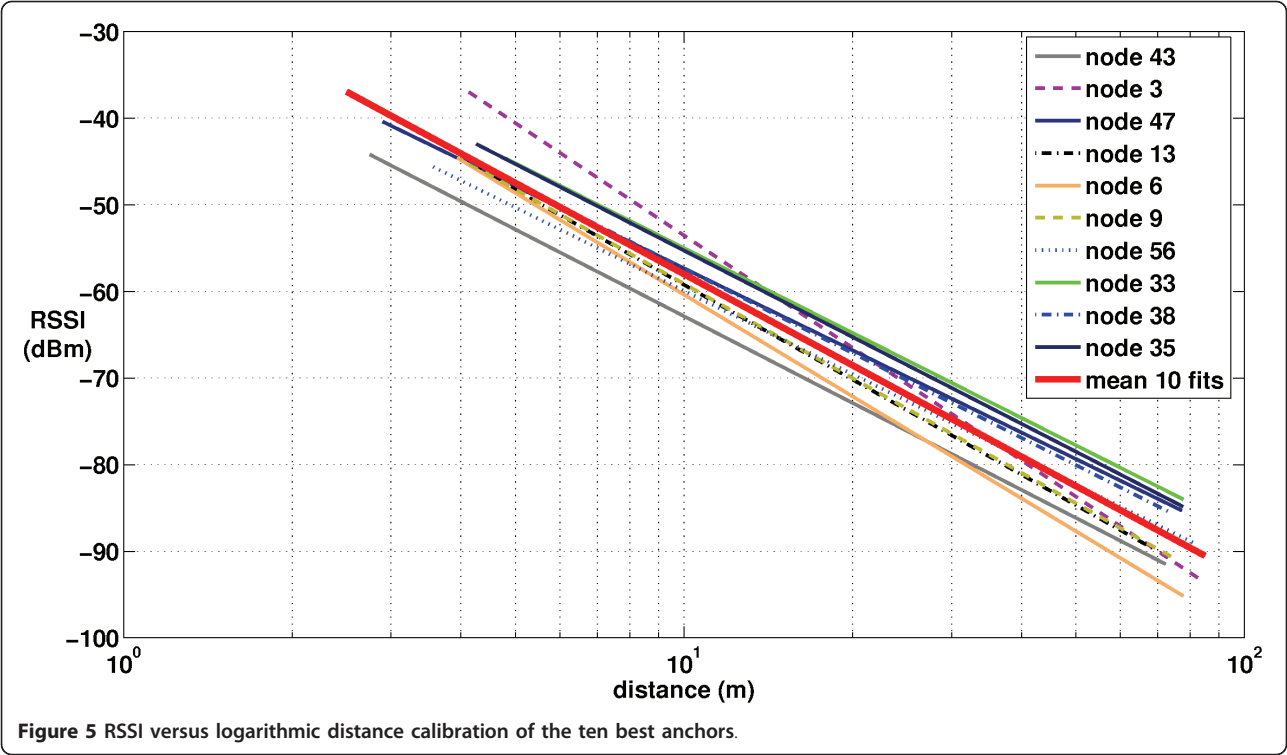
Procedure_Selection_of_Anchors_based_on_RSQ
Procedure_Selection_of_Anchors_based_on_Error_on_Distance
Procedure_Min-Max
If there are larger than expected large circles for the used hardware: delete them
Procedure_Ignoring_Circles
Delete the typical circles containing at least 5 others if there are no typical circles contained by at least 5 others
If there are typical circles contained by at least five others delete them.
Procedure_Maximum_Likelihood
Group the anchors based on distance
Are there small distance circles contained by other small distance circles of other group members?
If yes:
new typical circle radius all members = geometric average of $m$ measured distances
new_Error_on_Distance = (Error_on_Distance of member) /sqrt (# remaining typical circles -1)
adapt small and large circles according to new_Error_on_Distance
continue
If no: continue

receiving) with a linear regression tool: RSSI as a function of the logarithm of the distance (in m). RSQ is the square of the Pearson correlation coefficient [46]. “Zero” indicates that there is no linear fit, whereas a 1.0 indicates that all points lie on a straight line. For the RSSI at full power, the RSQ is varying from 0.89 on node 43 to 0.37

on node 21 as shown in Figure 4. The higher the RSQ, not only the higher the quality of the anchor nodes, but also the smaller their quantity. We suggest choosing RSQ of 0.8 (corresponding to a Pearson correlation coefficient of -0.89), resulting in 10 anchor nodes. The top 10 best-fitting nodes are marked with a square in Figure 2. They are in the extremities of the building, not in the corridors. These nodes are located in positions less sensitive to multipath fading and are considered as more informative ones. This is also confirmed by another empirical study in an indoor environment [47]. In Figure 5, the intercept points and the slope factors are utilized to produce graphs of the 10 best fits with  $RSQ > 0.80$ . A mean of these 10 best fits is also included. This line is constructed using the average of the 10 intercept points and the average of the 10 slopes as a new intercept point and slope, respectively. Figure 6 gives the reported RSSI versus the logarithmic distance with the best node 43 sending and all others receiving. It is known from Section 3 that there are 47 active nodes. Hence, in this half-duplex communication system there are 46 potential receiving nodes. In this figure, the measurements from nine receivers are below the sensitivity level. Hence, there are 46 minus 9, i.e., 37 measurement points. The graph also includes a linear regression of the measured RSSI and the mean of the 10 best fits. The last two lines are almost coinciding. Measured points above the fit are due



**Figure 4 RSQ correlation coefficient of RSSI versus logarithmic distance for the sending nodes on the third floor.** Some nodes are highly RSSI-log(distance) correlated.



certainly to (constructive) multipath fading as explained in the previous section, whereas the measured points below the fit are due to (destructive) multipath fading or attenuation. Figure 7 gives the worst-performing node 21 (which is in the longest corridor). For this central node, the measurements from only one receiver are below the sensitivity level, yielding 46 minus 1, i.e., 45 measurement points. The fit and the mean of the ten best fits are no longer parallel. The slope of this node is not so steep, multipath fading boosts the tail, and the higher attenuation at low distance cuts the head. Furthermore, there is a greater difference between the measurements and the fit.

Adding more nodes with lower RSQ will pile up with flatter slopes. Calibrating the ten best sending nodes with their respective intercept points and slopes will improve the RSSI to distance conversion. These intercept points and the slopes calculated with known distances can now be utilized to calculate any distance with a measured RSSI as input.

Hashemi [48] reports that many researchers had added an error correction to (1) and (2) and minimized it in their investigation. In [1,39,49], this is, e.g., done for the attenuation of walls. It is known from Section 3 that many walls of our building are plywood. Passing through them does not affect the RSSI measurements considerably. Likewise, propagation paths from the

sending nodes in the staircases will pass through the same highly attenuated concrete material for all the other nodes and therefore will mainly influence the intercept point. Since all anchors are individually calibrated, our model absorbs the latter effect. Although a model that includes walls can have a positive effect on the position accuracy, it will add complexity.

## 5.2. Selection of anchors with low standard error

With the ten best-correlated sending nodes, every node in the network hears at least 3 which should be sufficient for the 2D localization of a subject [50]. Some localization protocols work fine with anchor nodes in the corner [51]. Unfortunately, the building in this case (and many others) is not a square, and it is obvious that the absolute accuracy of far nodes will be lower. The challenge now is to find a few extra central nodes to increase this distance accuracy. Linear regression assumes a constant standard deviation for each point on the regression line [52]. There are two conditions for having a good RSQ. If the slope is flat and the points are spread around this line, then a low RSQ will result. On the other hand if the slope is not flat, but the points are spread too far around the line, then there will also be a bad correlation. The standard deviation is a measure indicating how close the points lie around the regression line and is, therefore, important for

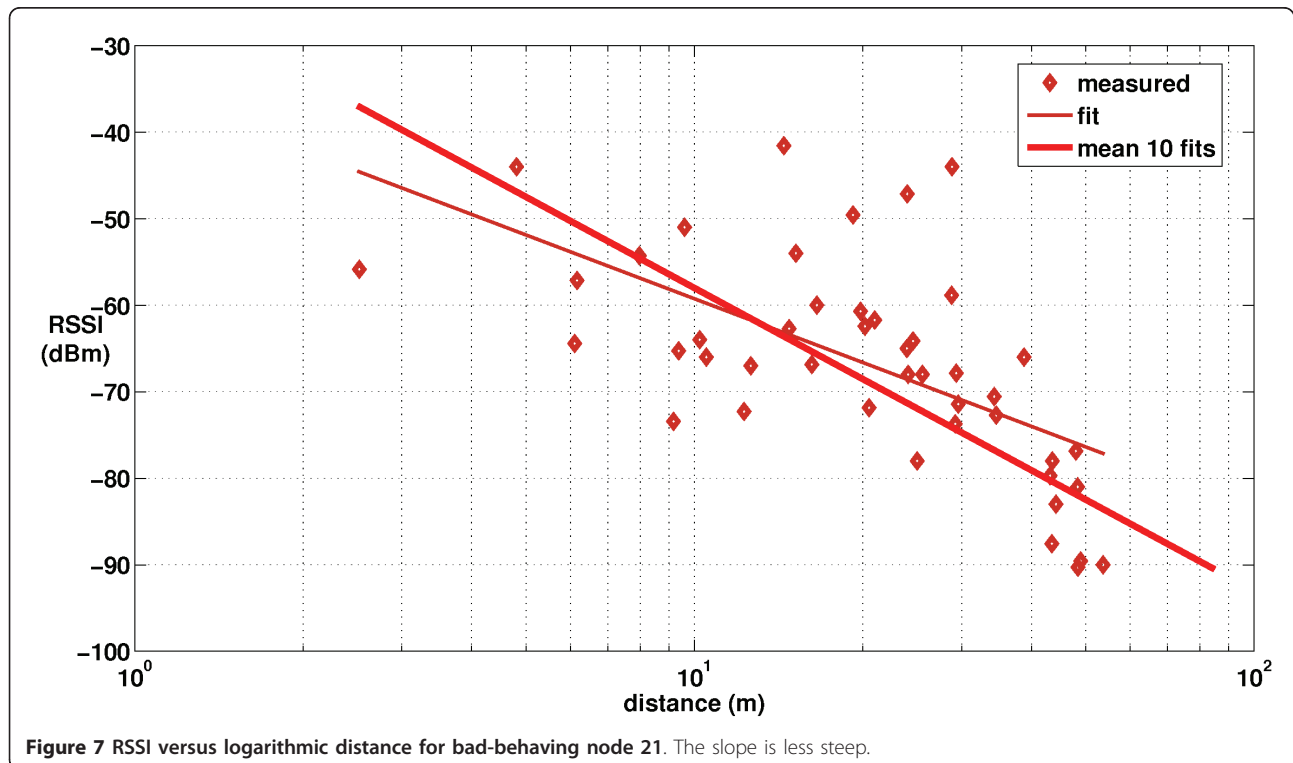


Figure 7 RSSI versus logarithmic distance for bad-behaving node 21. The slope is less steep.



determining the ranging bounds. These bounds can be theoretically determined using the Cramer-Rao algorithm, like in [53].

In this article, we propose a new more empirical approach for determining and optimizing the above bound. Indeed, any standard linear regression tool can calculate the standard error on the vertical axis which is an estimate of the standard deviation. We suggest swapping the RSSI and the  $\log(\text{distance})$  axes and redoing the linear regression, resulting in the standard error on the logarithm of the distance. We define the error on distance parameter as two times the estimated standard deviation of the logarithmic distance frequency distribution. It is noted that the RSQ is invariant to the axes swap: the RSQ values do not change when the axes are swapped.

In Figure 8, the frequency distribution is plotted for logarithmic distances of 1.3 (20 m), 1.4 (25 m), and 1.7 (50 m) with the same error on distance. It can easily be verified that with this constant error on distance, the absolute distance error in meter increases with increasing logarithmic distances. Therefore, central nodes, in addition to the ten anchor nodes in the extremities of the building, will improve absolute distance accuracy.

RSSI measurements that are too low (too much attenuation from the walls or too much destructive multipath fading) will result in distances at the right-hand side of the peak of the distribution and will thus correspond to estimated distances that are too high. RSSI

measurements that are too high (certainly too much constructive multipath fading) will result in distances that fall at the left-hand side of the peak of the distribution and will thus correspond to estimated distances that are too low. The above mentioned property is widely used in the design of our positioning algorithm in Section 6.

Figure 9 gives the error on distance for each node. Node 3 is the node with the lowest error on distance (0.21). It is known from Figure 4 that this is also the previously selected second-best anchor node based on correlation. As can be seen from Figure 10, where the RSQ is plotted versus the error on distance for each node, the nodes with the lowest errors correspond to the nodes with the highest RSQ, but there are also two new nodes in the center: node 30 and node 31. These nodes are marked with a circle in Figure 2. Inclusion of these nodes in our set of good nodes implies a lower absolute error on the position.

Measuring a logarithmic distance of  $1.5 \pm 0.3$  (31.6 m) will result in a distance of  $10^{1.2}$  to  $10^{1.8}$  or 15.8 to 63.1 m or 31.6 m (-50 to +100%). These error percentages are in a multipath environment an order of magnitude greater than the 10.1% encountered in less-complex environments [54], although equally well calibrated. It is also noted that the error percentages in plus and minus are asymmetric, because the distance on error is a logarithmic value, and thus the geometric mean needs to be considered.

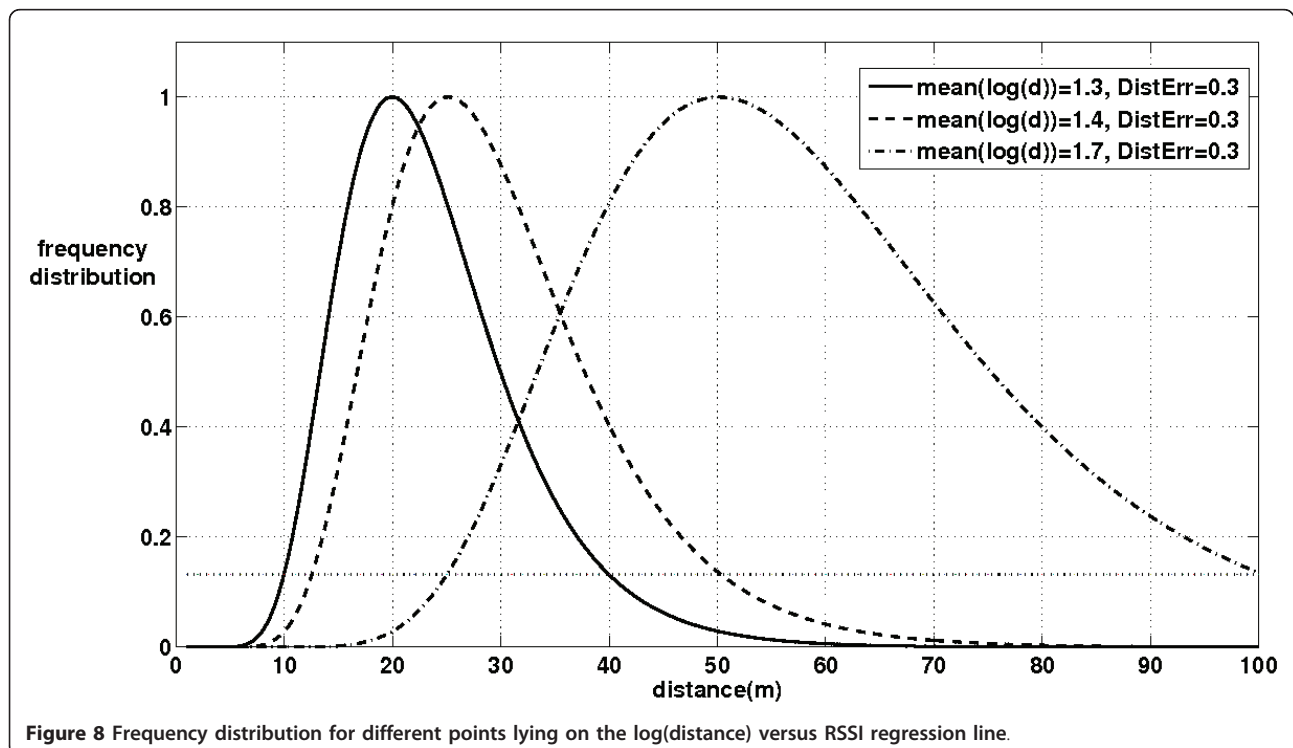
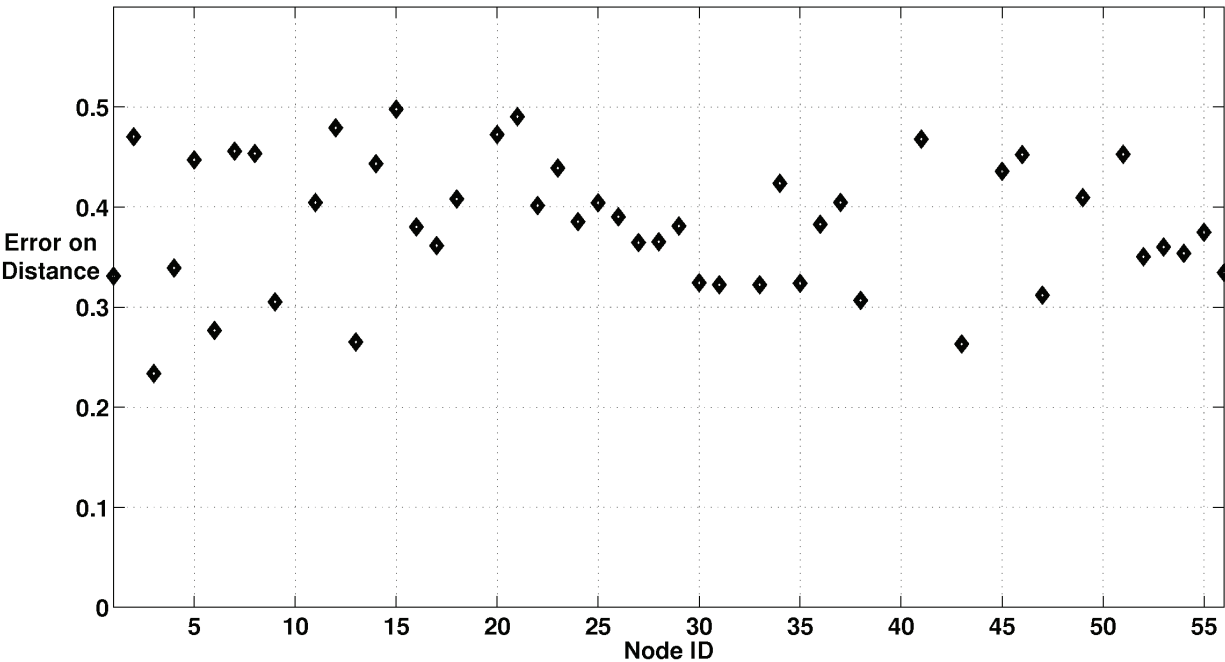
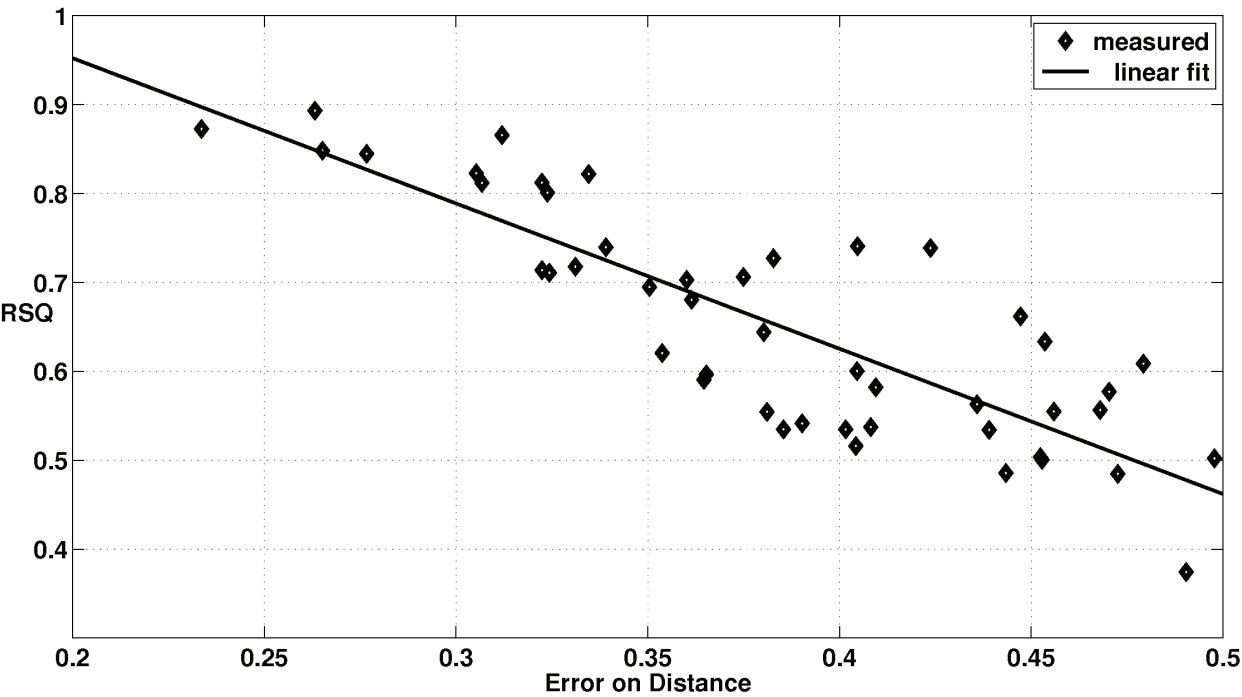


Figure 8 Frequency distribution for different points lying on the  $\log(\text{distance})$  versus RSSI regression line.



**Figure 9** Error on distance (defined as two times the estimated standard deviation of the logarithmic distance frequency distribution) for the sending nodes on the third floor. Some central nodes have low maximum error on distance.



**Figure 10** RSQ correlation coefficient versus error on distance showing a good linear fit.

We produce small and large distance circles with the error on distance and will use these from Section 5.4 onward. A typical circle corresponds to the calculated distance, obtained with the measured RSSI and the calculated slope and intercept for the corresponding anchor node. The radius of the small distance circle corresponds to the radius of the typical circle divided by  $10^{\text{error\_on\_distance}}$ . The radius of the large distance circle corresponds to the radius of the typical circle multiplied by  $10^{\text{error\_on\_distance}}$ .

In Figure 11, the regression is shown for node 30 with the swapped axes and two curves are added: the curves with the 95% expectation and 95% confidence intervals. These curves are hyperbola [52] with the average of the RSSI regression points as center  $X$ -coordinate and the average of the logarithm of the distance regression points as center  $Y$ -coordinate. If an individual logarithmic distance-RSSI experiment is performed, then 95% of the measurements will be found between the expectation lines. The area between the confidence lines (or confidence band) shows where the regression line is expected for the given confidence level. Around the center point, the intervals are the smallest. Toward both ends of the regression lines, the intervals become larger. This can be attributed to the fact that the uncertainty on the slope increases as the regression point is no longer in the center. Using this property gives an insight in the quality of our regression model: Anchor nodes at the extremities of the building have their maximum accuracy at the

geometric means of the minimum calculated distance (around 4 m) and the maximum calculated distance (around 83 m) or about 18 m. This is illustrated in Figure 12 with shaded rings for two anchors at the extremities of the building. At the geometric center of these rings, the accuracy of the regression model is maximal, since this corresponds to the central  $(X,Y)$  coordinates in Figure 11. Moving away from this central point in this figure results in wider confidence intervals resulting in a lighter gray shade as seen in Figure 12. The rectangle represents the building considered. Rings of central anchors will intersect with both the other rings and thus have a positive effect on the accuracy. These central anchors unfortunately cannot improve the accuracy of the model at the extremities of the building since there these are also working at the end of their regression lines.

### 5.3. Complexity of the calibration process and robustness against environmental changes

In the previous two sections, the best anchors are selected from the active nodes, and each anchor is calibrated with its own propagation constants (slope and intercept). This subsection discusses the complexity of such a calibration process and the robustness of the system against environmental changes.

After the experiment on the test bed the measurements (see Section 3) are imported in Matlab. For each sending-receiving pair, the RSSI is averaged and forms a RSSI-matrix, where element  $a_{i,j}$  corresponds to the

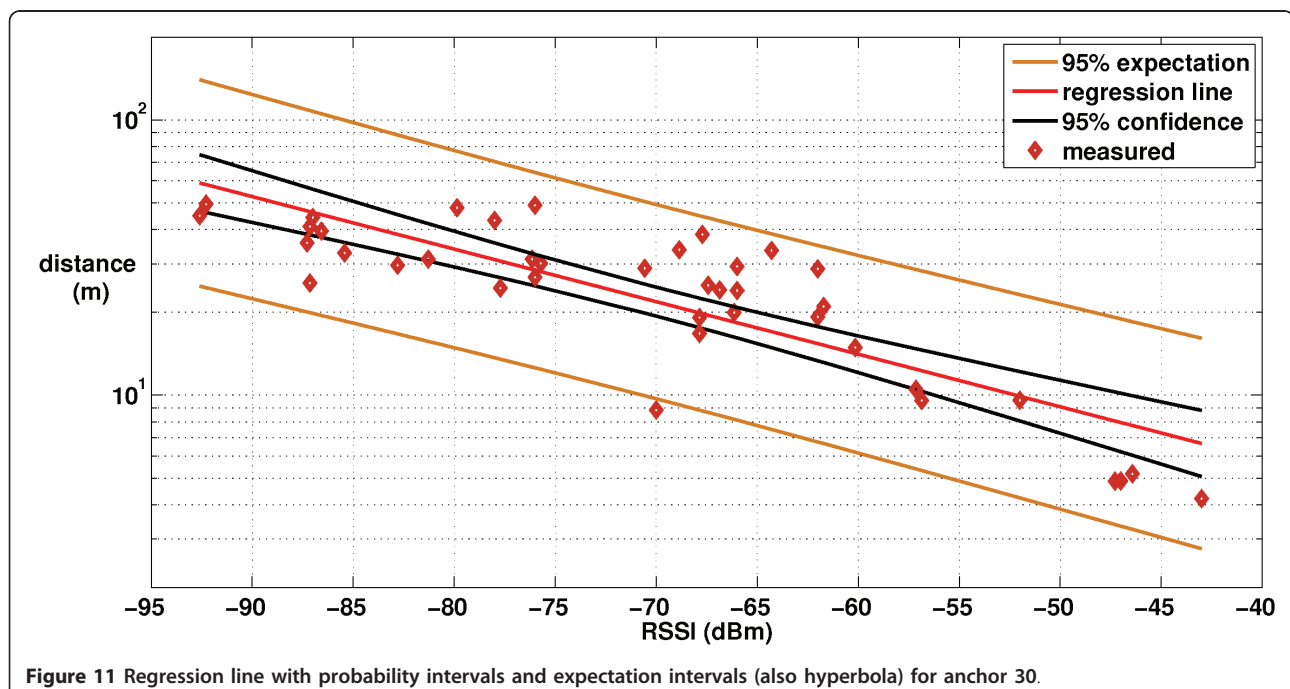
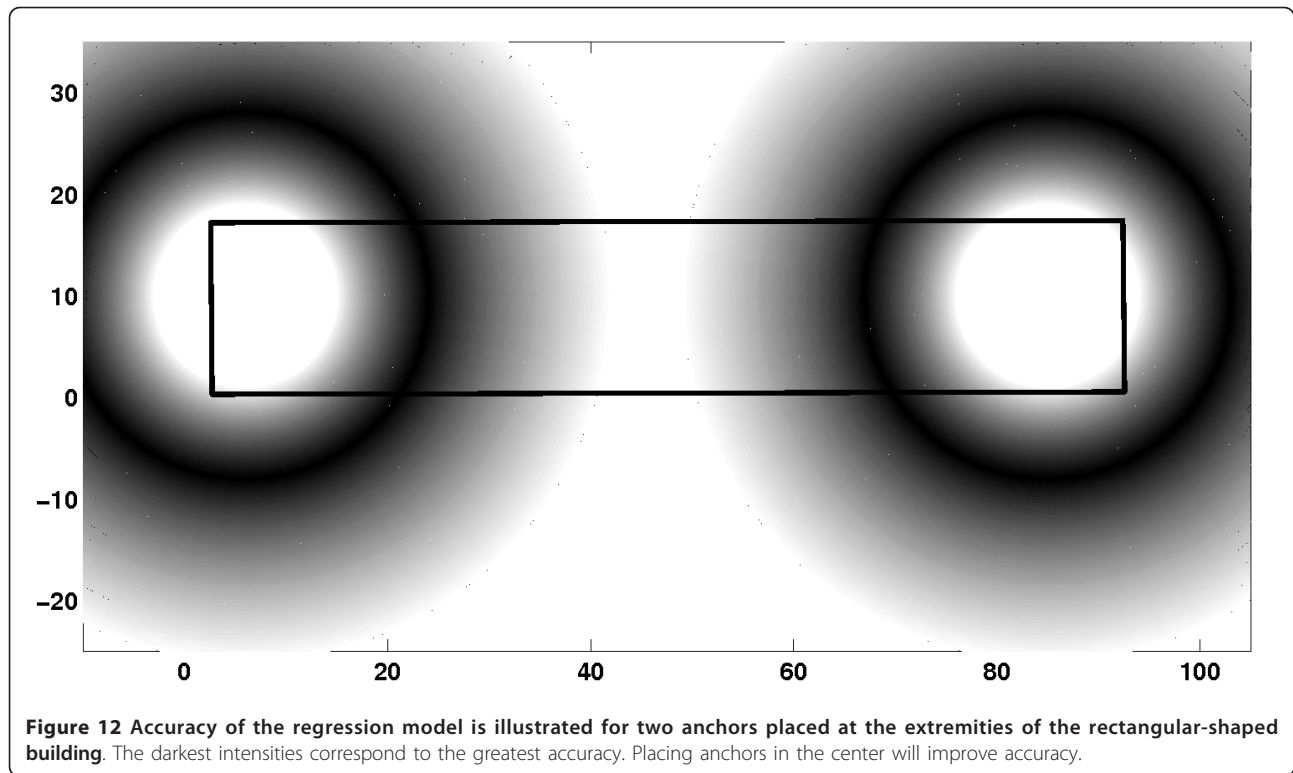


Figure 11 Regression line with probability intervals and expectation intervals (also hyperbola) for anchor 30.



averaged RSSI reported from receiver  $j$  with sending node  $i$ .

With the (known) position of the nodes, a distance matrix is formed too. Since our nodes are fixed this matrix remains the same for all experiments. On the corresponding rows of both matrices our software now calculates the slope, intercept, RSQ, and error on distance. It selects the ten best-correlated nodes and completes the selection with the two first nodes in the sorted error on distances vector, which are not in the best-correlated list yet. At this stage, the anchors are selected and calibrated. The location information of the remaining nodes is discarded in subsequent procedures, and the anchors are used to localize these nodes. Using a single power level and sending 240 packets (with a 25 ms inter-packet delay) over a single channel, the whole calibration phase is completed in less than 10 min, including the measurements. Calibration can easily be redone, e.g., when a node becomes inactive. Sending less packets increases the effect of time fading, since there are less-averaged RSSI measurements. Averaging multiple-channel RSSI measurements will reduce the effect of different reflections.

For a fingerprinting-based algorithm, a training phase is not only typical, but also essential. During this time-consuming phase (usually several days), a large database is filled with RSSI measurements. During the final phase, RSSI measurement is compared with all the

records of this database, and the record with the best match gives the estimated position. In this article, a different approach is used: the already deployed nodes in this study are utilized to quickly and automatically characterize the propagation constants (slope and intercept) of the medium. The knowledge of these two parameters is sufficient to localize (not previously stored) targets.

#### 5.4. Min-max algorithm

Our software now gathers data and produces overviews like Tables 2, 3, and 4. As an example, node 4 (Table 2), node 55 (Table 3), and node 54 (Table 4) are considered, but this is performed for all the targets.

In the first column, we find the emitting anchor, the second represents the reported RSSI by the target, the third gives the corresponding calculated distance (taking into account each anchor's own propagation parameters), the fourth the error on distance from Section 5.2, and the fifth the expected small and large distance, calculated based on the previous two columns. For example, for anchor 3 in Table 2, a calculated distance of 11.3 m with an error on distance of 0.234 gives an expected small distance of 11.3 times  $10^{-0.234}$  or 6.6 m and a maximum expected distance of 11.3 times  $10^{+0.234}$  or 19.4 m. The last column gives the RSQ from Section 5.1. Tables 5, 6, and 7 show the real distance between the anchors to node 4, 55, and 54, respectively.

**Table 2 Data available for positioning node 4**

Anchor ID	RSSI (dBm)	Calculated distance (m)	Error on distance	Expected distance (m)	RSQ
3	-55.9	11.3	0.23	6.6...19.4	0.87
6	-63.9	12.3	0.28	6.5...23.2	0.84
9	-59.3	10.1	0.31	5.0...20.4	0.82
13	-62.6	12.3	0.27	6.7...22.7	0.85
30	-78.0	35.6	0.32	16.9...75.2	0.71
31	-89.8	108.1	0.32	51.5...227.2	0.71
33	-90.4	122.2	0.32	58.2...256.7	0.81
47	-89.6	105.7	0.31	51.2...216.8	0.87

Section 5.1 found ten good anchors, having a good linear regression between the RSSI reported by all receivers and the logarithmic distances between sender and receiver. Section 5.2 added two good central anchors having a low error on distance. Furthermore, it revealed where the anchors are the most accurate. In this section, we start polishing up the measurements.

When comparing Table 2 with Table 5, a large error between the calculated (122.2 m) and the exact distances (72.0 m) from anchor 33 is seen. Although not an outlier here (58.2 m is less than 72.0 m), this measurement can disturb any positioning algorithm. More extreme examples are found with other targets with calculated distance up to 134 m (1.5 times the size of the building) and maximum distance up to 275 m (three times the size). Multipath fading resulted in flatter slopes of the regression lines, and the extrapolation of these lines causes these large errors.

Similar to [55], a combination of a min-max algorithm with a maximum likelihood algorithm is needed to improve these situations. Contrary to [55], we suggest dropping all the measurements where the calculated distance is greater than a limit because the correction to that limit can lead to a large error. Based on Equation 2, a maximum distance for this hardware is calculated. Indeed, when both a sending power level of 0 dBm and a receiver sensitivity of -92 dBm (2) are considered, a maximum distance of 82.8 m is obtained. Thirty-five individual measurements are dropped in this way. For

all the nodes, this additional step improves the distance error. If a distance calculation is higher than the limits for the hardware, then it probably comes from a connection with either too much attenuation or too much destructive multipath fading. In both cases, it comes from anchors operating at the end of their regression line. This algorithm when applied to node 4 (in Table 2) results in the elimination of the RSSI measurement of anchors 31, 33, and 47.

### 5.5. Elimination of circles

The elimination of circles discussed in this section is based on the accuracy of the model of Section 5.2. Anchor 54 is representative for a target in the center of the building, and anchor 55 is a good example of a target in an extremity of the building. These nodes perfectly demonstrate the accuracy of the model.

With column 3 of Tables 2, 3, and 4 and the exact position of the anchors, typical (dot marked) circles can be drawn for locating the nodes. This is done in Figure 13 for node 4. The radius of these typical circles is the calculated distance (based on RSSI measurements) to the target, represented by the dark square. If all measurements were on the regression line, all typical circles would intersect in the target. The too large (red) typical circles of anchors 31, 33, and 47 contain all the other typical circles. The RSSI measurement is too low, and in the  $\log(\text{distance}) - \text{RSSI}$  graph, the measurement is too

**Table 3 Data available for positioning node 55**

Anchor ID	RSSI (dBm)	Calculated distance (m)	Error on distance	Expected distance (m)	RSQ
13	-89.3	66.9	0.27	36.3...123.2	0.85
30	-68.9	20.2	0.32	9.6...42.7	0.71
31	-74.0	38.8	0.32	18.5...81.5	0.71
33	-44.0	4.6	0.32	2.2...9.7	0.81
35	-66.3	21.4	0.33	10.2...45.2	0.80
38	-61.9	13.7	0.31	6.8...27.8	0.81
43	-62.4	9.7	0.26	5.3...17.8	0.89
47	-58.3	10.7	0.31	5.2...22.0	0.87
56	-56.3	7.7	0.33	3.6...16.6	0.82



**Table 4 Data available for positioning node 54**

Anchor ID	RSSI (dBm)	Calculated distance (m)	Error on distance	Expected distance (m)	RSQ
3	-84.1	51.3	0.23	29.9...87.8	0.87
6	-85.1	43.0	0.28	22.8...81.4	0.84
9	-88.9	66.1	0.31	32.7...133.6	0.82
13	-77.1	31.0	0.27	16.9...57.1	0.85
30	-70.0	21.7	0.32	10.3...45.8	0.71
31	-46.0	6.3	0.32	3.0...13.3	0.71
33	-73.4	36.8	0.32	17.5...77.3	0.81
35	-60.0	13.9	0.33	6.6...29.2	0.80
38	-54.4	8.1	0.31	4.0...16.4	0.81
43	-70.0	16.4	0.26	8.9...30.0	0.89
47	-73.6	32.8	0.31	16.0...67.3	0.87
56	-73.6	26.7	0.33	12.4...57.7	0.82

far above the linear regression line. As explained previously, these circles have been deleted.

With two measurements (from two different anchors) of the same target, there are four possibilities: they are both too small, they are both (slightly or much) too large, the first one is (slightly or much) too small and the second is (slightly or much) too large or vice versa. On averaging the first two cases we get no improvement, since the closest distance to the target is corrected in the wrong direction. On averaging in the last two cases, only there is an improvement if there are no outliers. Deleting bad measurements and outliers is a major preprocessing step in the case of the heavily multipath faded environment, discussed in this article.

We suggest expanding the elimination of bad measurements based on containing-contained circles and the quality of the anchors. Figure 14 shows the typical circles for node 55 constructed using Table 3 and the exact position of the anchors. We observe that the largest (yellow) typical circle (belonging to anchor 35) contains five others. The smallest (pale blue) circle at the

left-hand side (belonging to anchor 33) is contained by three others. Node 55 is representative for all the nodes at the extremities of the building: there are no circles that are contained by minimum five others, and there are more than five containing circles. For targets at the left-hand side extremity of the building, all anchors are operating at the end of their linear regression line. The left-hand-sided typical circles are at the minimum distance, and all others (including the central) at the maximum distance of their respective regression line. The accuracy of the smallest typical circles increases when the target shifts to the center, thus these circles are not deleted. Between the large amount of large typical circles, there exists not only inaccuracy but also redundancy. In this situation, all typical circles containing at least five others will be ignored.

This number five is dependent on the total number of sending anchors (received by the target). We suggest using half of the total number of anchors minus 1. When this figure is set too low, good measurements may be eliminated. This will imply that the number of

**Table 5 Real distance between target node 4 and the 12 anchors**

Anchor ID	Real distance (m)
3	13.0
6	12.2
9	12.6
13	15.4
30	43.1
31	43.4
33	72.0
35	72.9
38	68.2
43	67.3
47	72.2
56	75.8

**Table 6 Real distance between target node 55 and the 12 anchors**

Anchor ID	Real distance (m)
3	82.5
6	77.8
9	73.9
13	68.4
30	33.7
31	33.9
33	4.8
35	12.9
38	15.4
43	10.9
47	6.8
56	5.0

**Table 7 Real distance between target node 54 and the 12 anchors**

Anchor ID	Real distance (m)
3	52.7
6	47.9
9	43.9
13	38.4
30	8.8
31	5.6
33	25.5
35	24.7
38	20.0
43	19.6
47	24.5
56	28.1

measurements is not large enough for the (good) positioning. When it is set too high, bad measurements may be allowed. This will have a negative impact on the next step of our preprocessing: the maximum likelihood algorithm.

Likewise, for targets in the center of the building, the anchors at the extremities are not extrapolating their regression lines, and thus, the largest circles should not be deleted now. The anchors in the center of the building are operating at low distances. In this case, too small central typical circles are contained by many

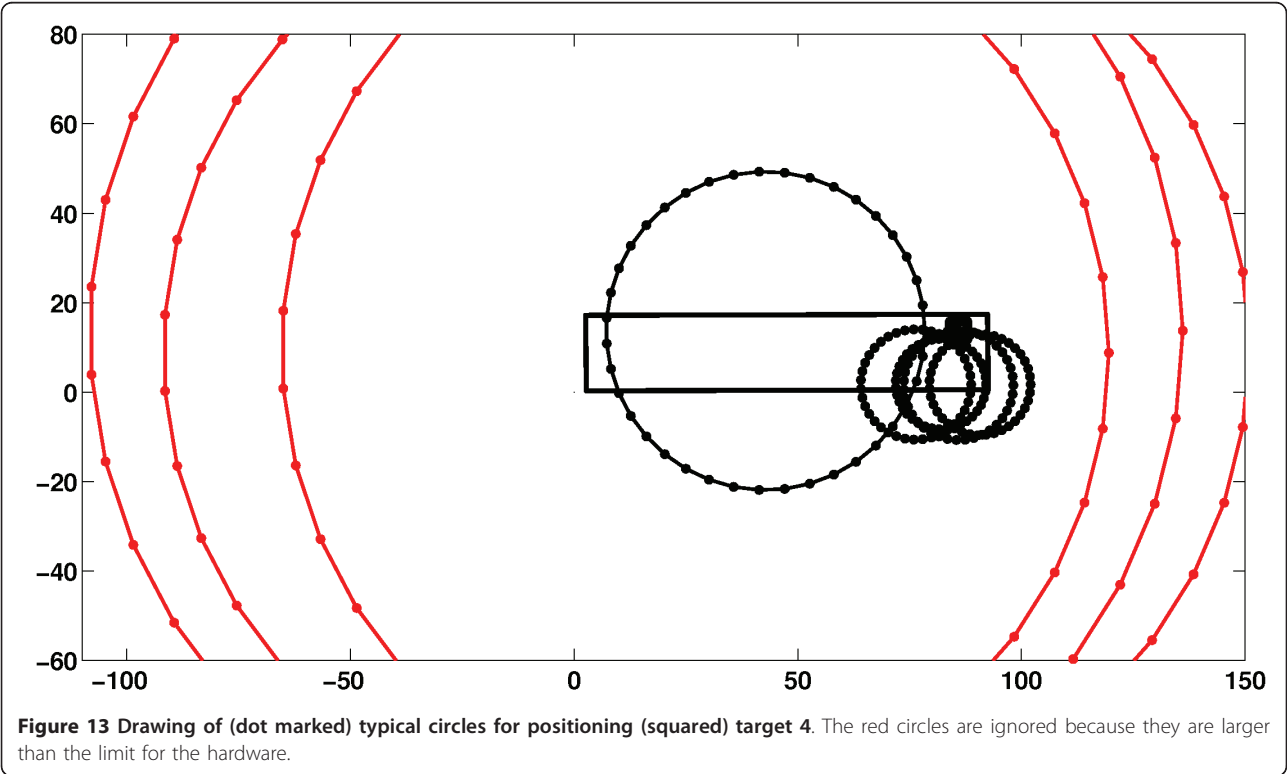
typical circles (coming from anchors at both sides of the extremities). A representative example is node 54. A closer look at Figure 15 reveals that the green typical circles contain minimum five other typical circles, and the blue circle is contained by at least five other typical circles. Hence, when there are both minimum five containing and five contained typical circles, only circles that are too small (blue) are deleted.

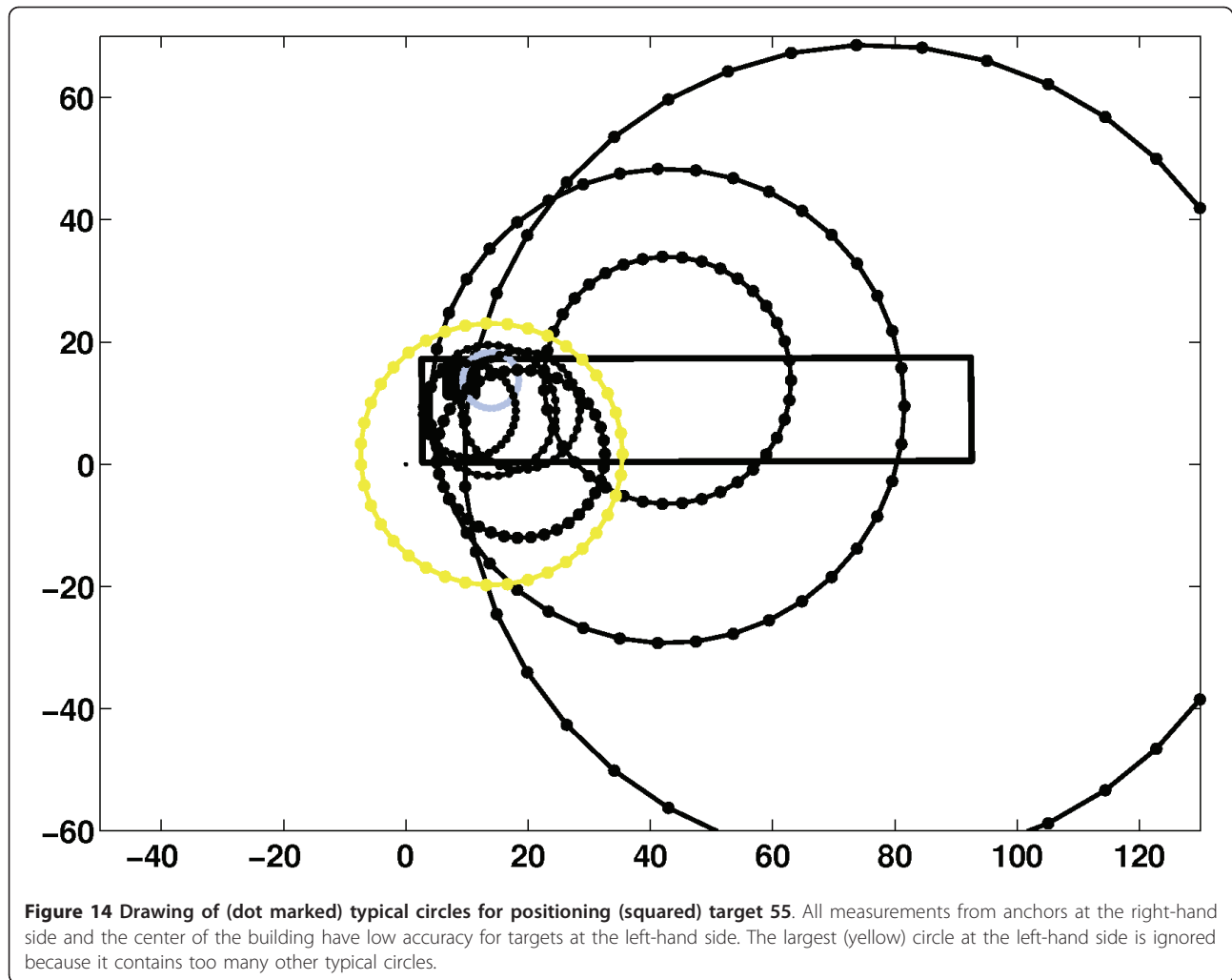
It is noted that the remaining typical circles are more consistent. The typical circles that are too large (green) will be used in the maximum likelihood algorithm on the distance of the next section, where they can correct the distances of their group members.

5.6. Grouping anchors

When anchors are grouped, a maximum likelihood algorithm can be used. Usually this algorithm is found in the positioning algorithm itself, but we will use this technique in our preprocessing to accomplish a better distance estimation. In Section 7.2, we will compare the results of our complete localization algorithm with the more conventional maximum likelihood on the position.

Figure 2 shows three groups of anchors: nodes 3, 6, 9, and 13 in the first group; nodes 30 and 31 in the second; and the others in the third. The nodes in each group are neighboring anchors in one extremity of the building, the center and the extremity of the building at the other side, respectively. We suggest considering the



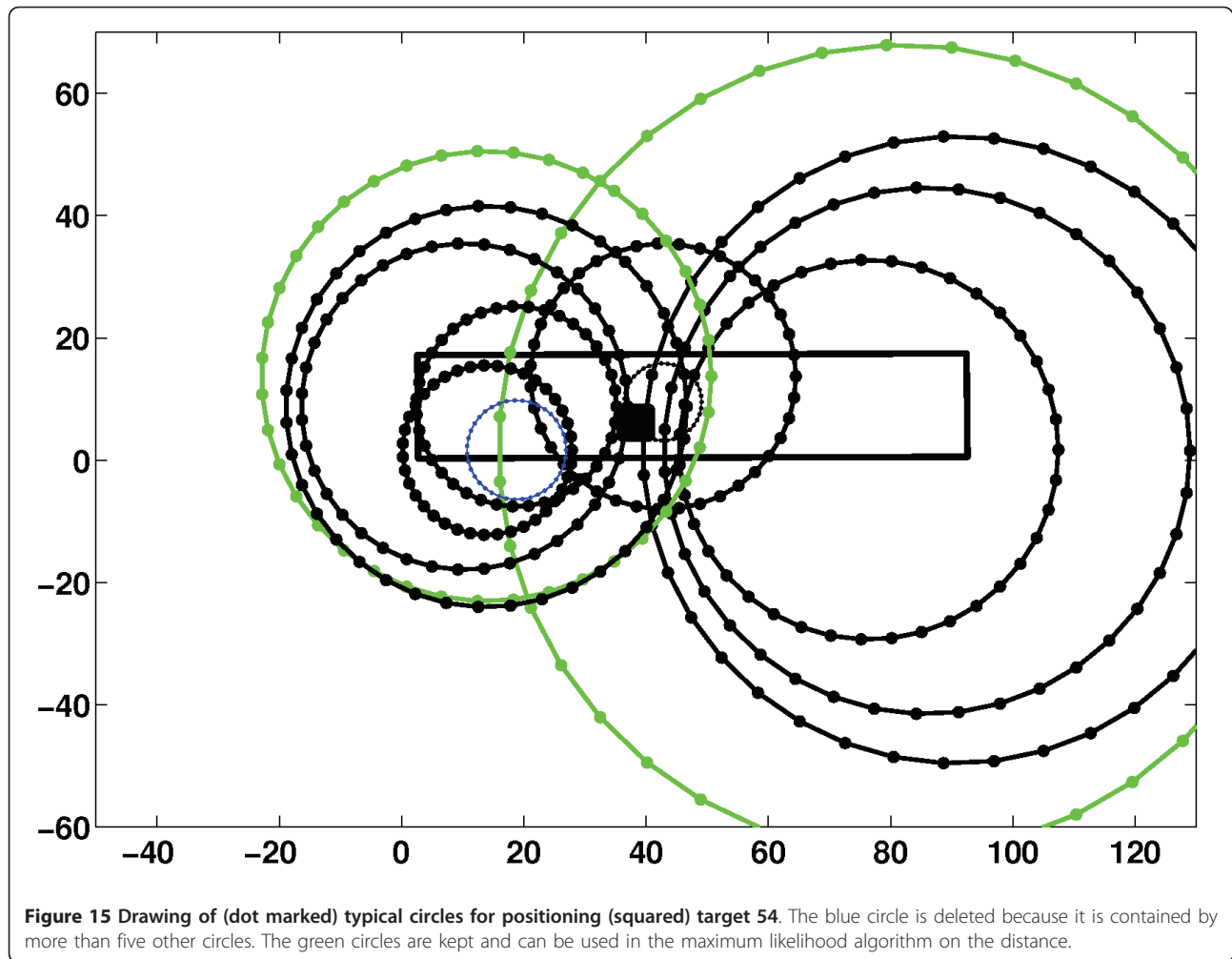


anchors in one group as coinciding if one small circle is contained in the other(s). It is known from Section 5.2 that the radii of these small circles are available from the fifth column of Tables 2, 3, and 4 and calculated using the error on distance.

Under the assumptions that the probability distribution of the logarithmic distance is normally distributed for all the group members with the same variance, the likelihood function equals the product of these normal distribution functions of the  $m$  (see Table 1) remaining measured logarithmic distances in each group. Partial derivation of this likelihood function [52] proves the most likely logarithmic distance equals the average of the measured distance of the  $m$  group members. Hence, the most probable distance can be determined based on the (geometric) average of the  $m$  (see Table 1) remaining measured distances in each group. This geometric average is calculated and applied to all anchors in that group. (It is noted that the position of the anchors remains unchanged.) If a distance estimation of one

anchor is too high, then it is compensated by the other group members. With distances too low also this is the case. By eliminating the outliers before this process, the chance of decreasing the accuracy of the good anchors is minimized. If the small circles do not contain others, then the target is probably nearby and the individual anchors are needed to increase the accuracy. In a more conventional study, the standard error for use in a maximum likelihood algorithm is determined using the Cramer-Rao Bound or calculated with a mean square error applied on the group members. In this article, we suggest using the standard error of the regression model and divide the error on distances by the square root of the number ( $m - 1$ ) to obtain the corrected error on distance for the group members. This method provides larger large circles and smaller small circles. The advantage of doing so will be explained in the next section.

In Figure 16, the maximum likelihood algorithm on the distance is illustrated. All the solid circles are small

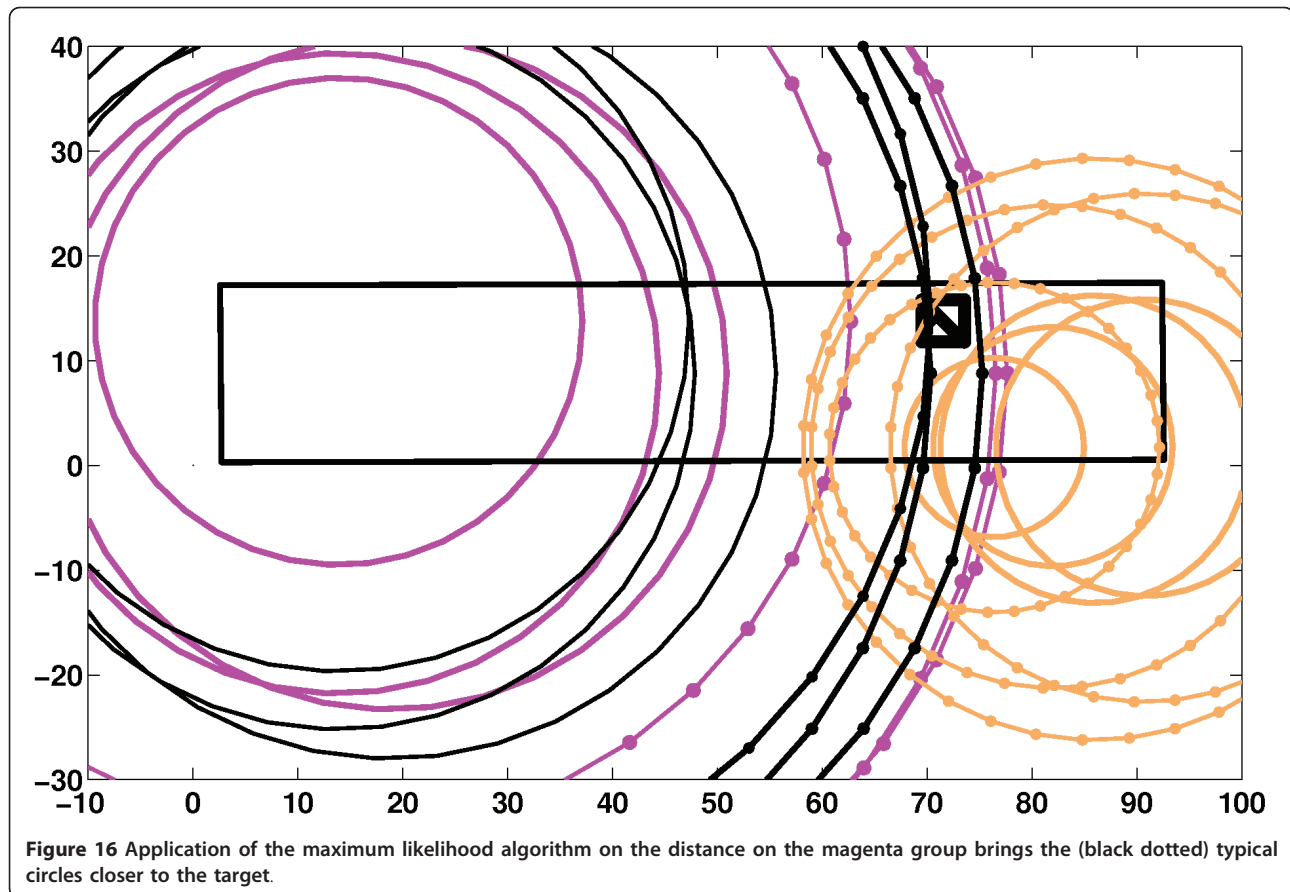


distance circles. Again, the typical distance circles are dot marked, and the target (node 14) is represented by a small square. The orange circles are for the first group with the anchors 3, 6, 9 and 13. In this case, none of the small circles is contained by another small circle of any group member. Therefore, the maximum likelihood on the distance is not applied. The magenta circles belong to the third group, consisting of anchors 33, 35, 38, 43, 47, and 56. It is noted that the measurements of anchors 35 and 56 are below the noise floor, and the measurement of node 38 is above the limits of the hardware (Section 5.4). The remaining (magenta dot marked) typical circles are not close to the target. Since the small distance circle of node 33 is contained by the small distance circle of nodes 43 and 47, the maximum likelihood algorithm on the distance is applied. The corresponding radii are calculated using this section and result in the black circles. The magenta distance circles are replaced by the black. It is noted that the corrected typical circles are now very close to the target. It is also noted that the large circles and the circles of second

group consisting of anchors 30 and 31 are not shown in this figure. Their inclusion would have overloaded this drawing.

## 6 Positioning algorithm

With the preprocessing of the previous section good input is achieved for positioning: Good anchors are selected, based on high RSQ and low error on distance; bad measurements are eliminated because both too small and too large distance circles are rejected. Furthermore, the grouping of anchors allows a maximum likelihood algorithm on the distance. All of this results in more consistent distance circles. Our error on distance approach results in small and large distance circles forming rings (washers) with accurate bounds on the distance. These distance rings can be used in an area-based localization algorithm as done in [56]. In that more theoretical study, those authors found both the intersection of the rings and the smallest enclosing circle covering this intersection of rings. It is noted that our algorithm provides a method to empirically obtain



these rings. The thickness of the rings is not only anchor dependent (see Tables 2, 4, and 6) but (because of the logarithmic property of the error on distance) also distance dependent (see Figure 8). For this approach, however, it is necessary that the rings are not disjunct. This is why the bounds of Section 5.6 need not be too conservative. Indeed, decreasing the thickness of the rings will increase the chance of these rings to be disjunct. For a few cases, this condition is not met in our multipath environment. Instead of increasing the bounds, a low complexity point-based algorithm better suits our empirical data. We present the flowchart in Figure 17. It mainly consists of six procedures and two decisions. The numbers represent the subsections where the respective items are discussed.

The main path is leading through 6.1, 6.2, 6.3, 6.4, 6.5, and 6.7. In a very few cases, procedure 6.6 is encountered, and the use of procedure 6.8 is even rarer.

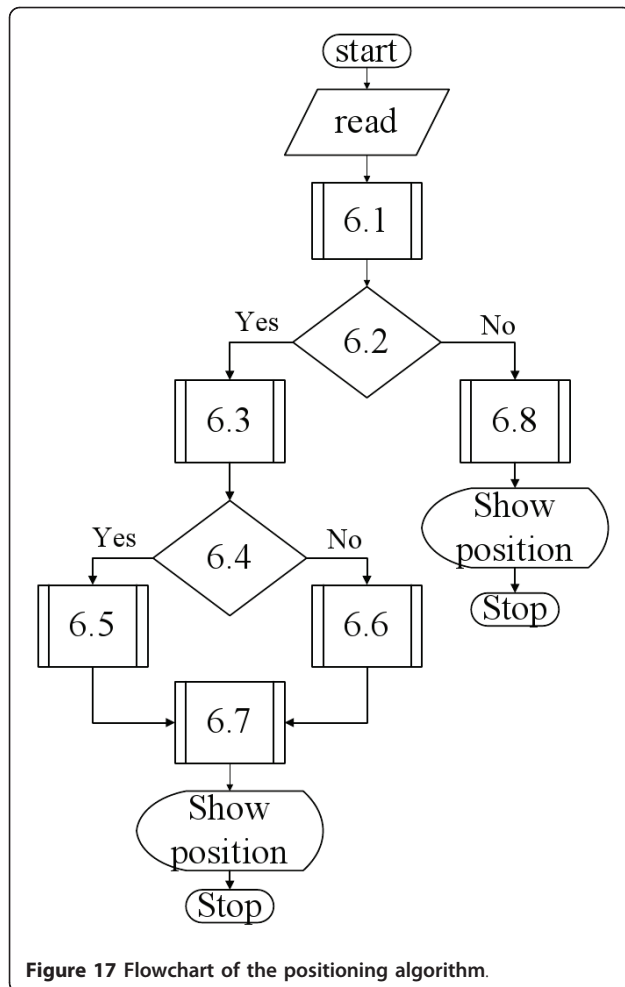
#### 6.1. Calculate intersection points of the longitudinal borderlines of the building and the large circles

For each target, the previous section resulted in small, typical, and large distance circles around the anchors.

This starting procedure looks for the common intersection of these large circles and the longitudinal borderlines of the building, see Figure 18. This figure is meant to illustrate the principle, and the redundant data, such as target number, anchor IDs, etc., is left out, to avoid overloading the drawing. This approach is valid for the remainder of this section. The rectangle represents the building, and the large circles are denoted by dashed lines. The intersection points of the large circles and the longitudinal borderlines of the building are calculated. If the leftmost intersection is to the right of the corresponding leftmost corner of the building, then it replaces this corner point. Now, the next large circle is considered, and its leftmost intersection point replaces the previous if it is more to the right. In this way, the left points are shifted from point A through D. The interpretation of this point is straightforward: if there are no outliers, then all the large circles represent effective maximum bounds on the distance, and the target cannot be at the left of the most restrictive left point.

This is adopted for both the longitudinal borderlines of the building. An analogous procedure gives





the leftmost right points of a trapeze. Without outliers on the constructive multipath fading, the large circles will not be too small, and the vast majority of the targets will continue to move on the main track toward 6.3.

### 6.2. Is the amount of constructive multipath fading acceptable?

As discussed in Section 4, constructive multipath fading results in too large RSSI measurements and hence too small distance circles. If the most restrictive large circles do not intersect, then constructive multipath fading is present (see Figure 19). The final right point is now situated left to the final left point. The trapeze now has “negative” sides.

When no outliers on the constructive multipath fading are present as shown in Figure 18, the trapeze has positive sides. Hence, this is a fast method to decide on the amount of constructive multipath fading. On the positive trapeze sides outcome, the algorithm continues with 6.3, and with negative sides 6.8 is the next step.

### 6.3. Calculate typical circles intersection points within the building

Next, the typical circles intersection points within the building are calculated. There are two possibilities: Figure 20 shows the situation where the intersection points are inside the building, and Figure 21 illustrates an example where typical circles do not intersect or have intersection points outside the building. Our software counts the intersection points within the building of the most intersected typical circle(s). In the next step, a new decision needs to be taken.

### 6.4. Is there more than one intersection point within the building?

Preprocessing steps resulted in more consistent distance circles. As discussed in Section 5.1, constructive multipath results in too small circles. When all the typical circles do not intersect at all within the building or when there is only one intersection point of all intersecting typical circles within the building, uncorrected constructive multipath fading is expected. This results in the selection of a different initial point. Very few targets follow the 6.6 procedure, and the vast majority of targets will continue with procedure 6.5.

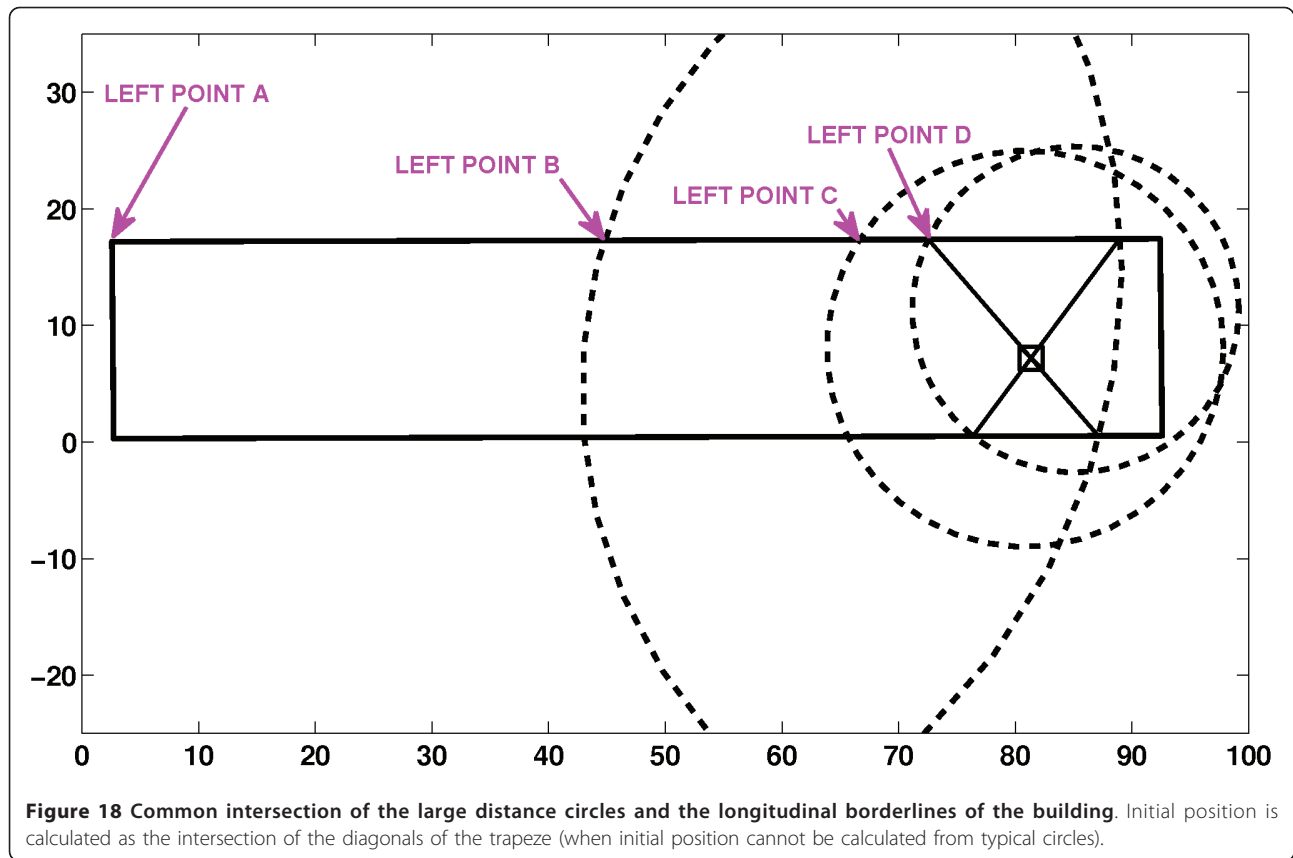
### 6.5. Calculate initial position as the centroid of all typical circles’ intersection points

A logical next step is the calculation of an initial position.

It is the result of a multilateration process, calculated as the centroid of all typical circles’ intersection points with the typical circle that is intersected most. Figure 20 gives an example where the leftmost circle is intersected twice. The little square in the middle of the drawing represents the initial position. In a few cases a typical circle is intersected up to 14 times by other typical circles.

### 6.6. Calculate the initial position as the intersection of the diagonals of the trapeze

In Figure 21, the leftmost typical circle is smaller than the corresponding circle of the same anchor in Figure 20. It is likely that there is more constructive multipath fading for Figure 21. The typical circles do not intersect within the building, either because there are no intersections or because the intersections are not within the building. It is also noted that if the multipath fading for the two rightmost anchors decreases (and these circles become larger), then there will again be intersections within the building. Therefore the constructive multipath is expected when there are no intersections within the building. Furthermore, when there is at the most one intersection point of all the typical circles within the building, the initial point cannot be calculated with



the multilateration step of Section 6.5. In this case the initial position is estimated as the intersection point of the diagonal of the (positive) sided trapeze, calculated from the large circles as shown in Figure 18. The procedure described in Section 6.7 will further increase accuracy.

#### 6.7. Calculate the estimated position of the nearest intersection points of the longitudinal line and the typical circles

Around the initial point, accurate typical circles are expected, since the preprocessing resulted in consistent circles. It is known from Section 5 that all typical circles would intersect in the target if there were no multipath fading. Furthermore, Figure 11 shows that the 95% confidence intervals (in m) are the widest at high distances. An examination in the neighborhood of the initial point at distances larger than the limits of the confidence interval will include outliers. Therefore, we impose an absolute limit on the search region around this initial point. When set too high, non-detected outliers will spoil the good measurements, and when set too low, good measurements from far anchors will be eliminated. As a limit, we propose the same total confidence interval of 22 m for all the anchors. Models that are more

complex can adapt this margin to the considered anchor. Figure 11 shows that for node 30 a 95% confidence interval of 22 m is reached at a distance of 50 m.

Figure 22 represents the final step in the localization. Around the initial position (obtained with procedure 6.6), a line of  $\pm 11$  m is drawn in the longitudinal direction. It is noted that the limits of the line are symmetrical, while the confidence intervals are not. Models that are more complex could keep these limits asymmetrical. Now, for each typical circle, the intersections are calculated with this line. The point that is the closest to the initial is kept. Multilateration of these points results in the estimated position. Owing to the form factor of the building, multilateration in the lateral direction of the building is not performed.

#### 6.8. Calculate the position as the intersection of the diagonals of the trapeze

This procedure is executed when at least one of the concerned anchor-target paths exhibits too much constructive multipath fading. In our building the preprocessing step reduced the number of the affected targets from 5 to 3. Instead of finding out which circle is too small, we allow negative-sided trapezes. Considering that both circles are equally likely to be too small and

that the negative sides are small, the position is estimated as the intersection of the diagonals of the trapeze. In Figure 19, this position is marked with a black square.

## 7. Results

In the first subsection of this section, the distance error is discussed, whereas in the next section, position errors are treated. The distance error only considers one of the 1942 sender/receiver pairs at a time. The error between the real and the calculated distance results in one point in the cumulative distribution plot (cdf). For the position error, all the distances from the different anchors are grouped, resulting in one of the 47 target positions at a time.

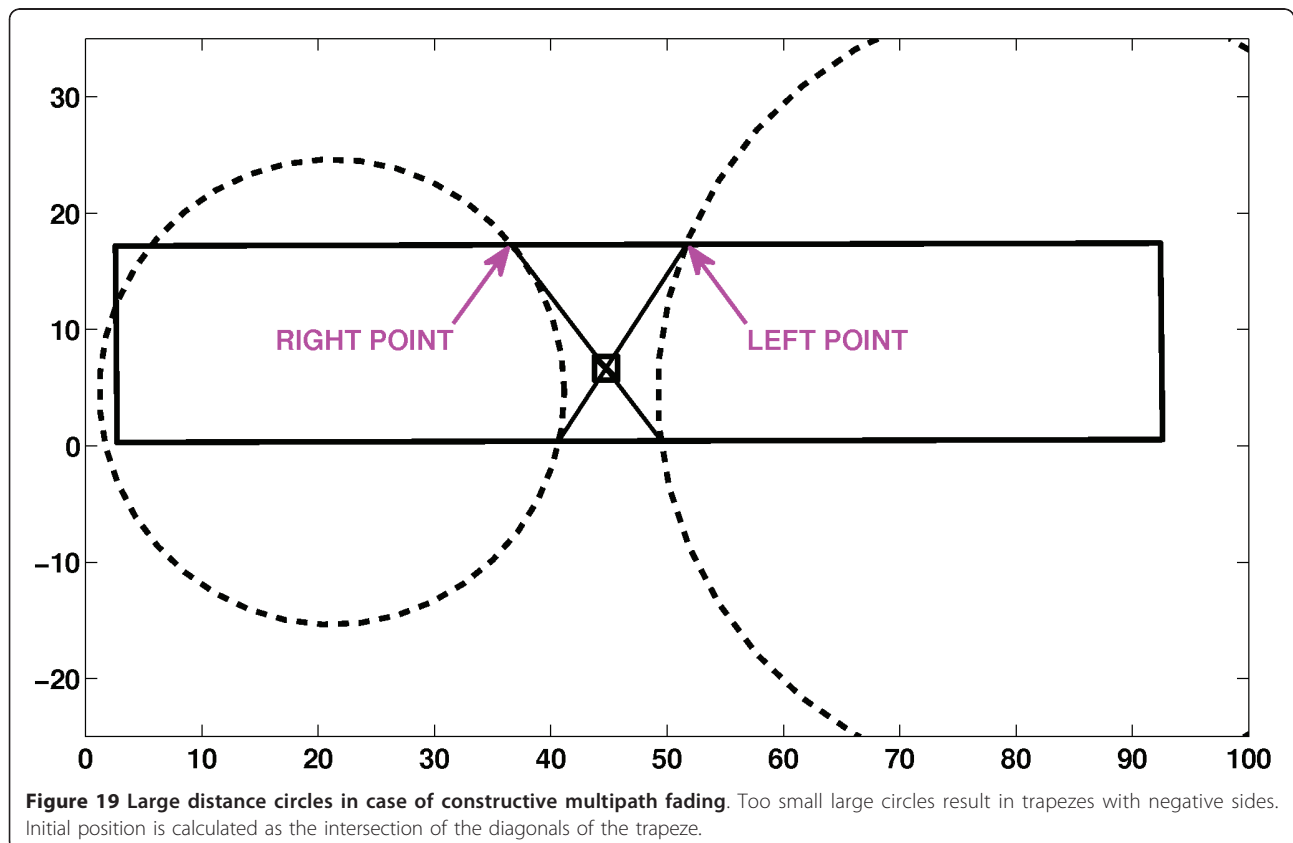
### 7.1. Distance error

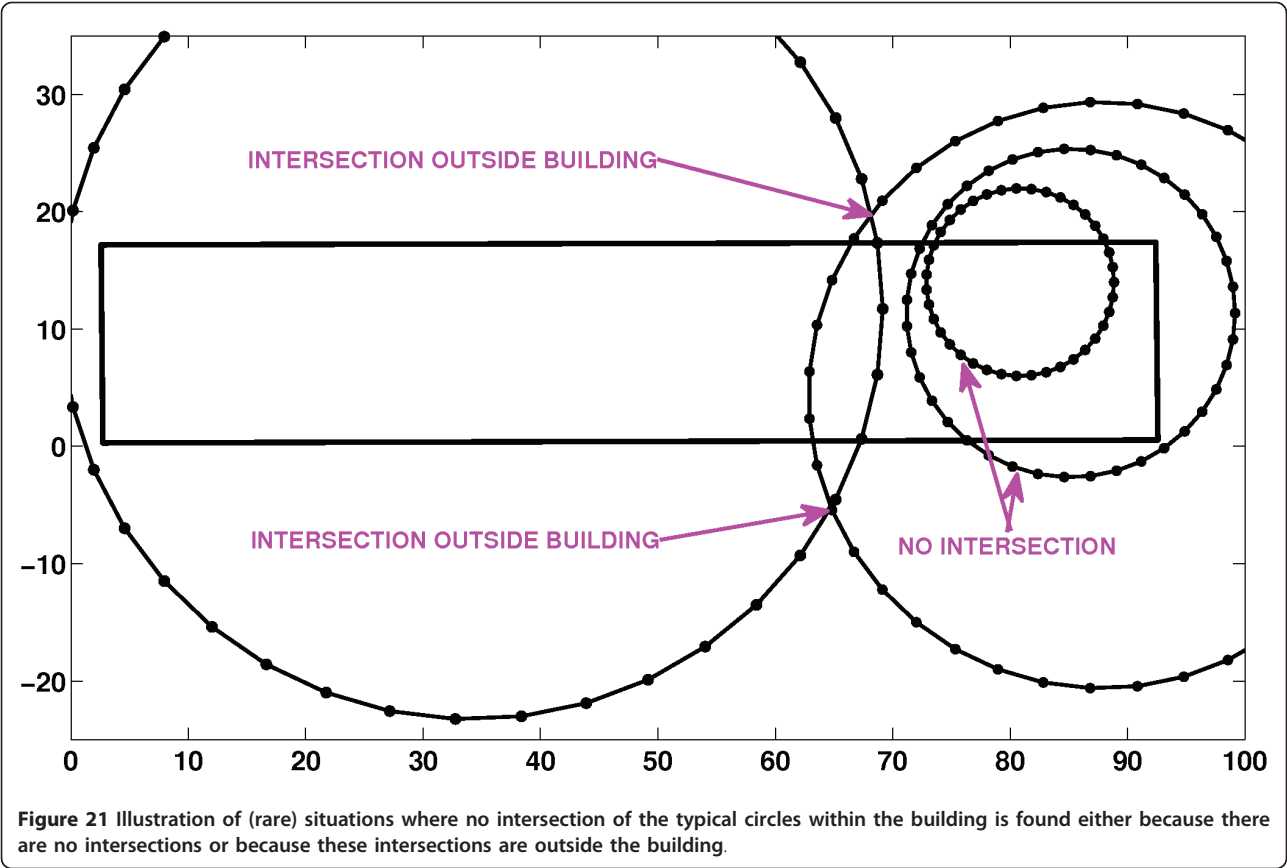
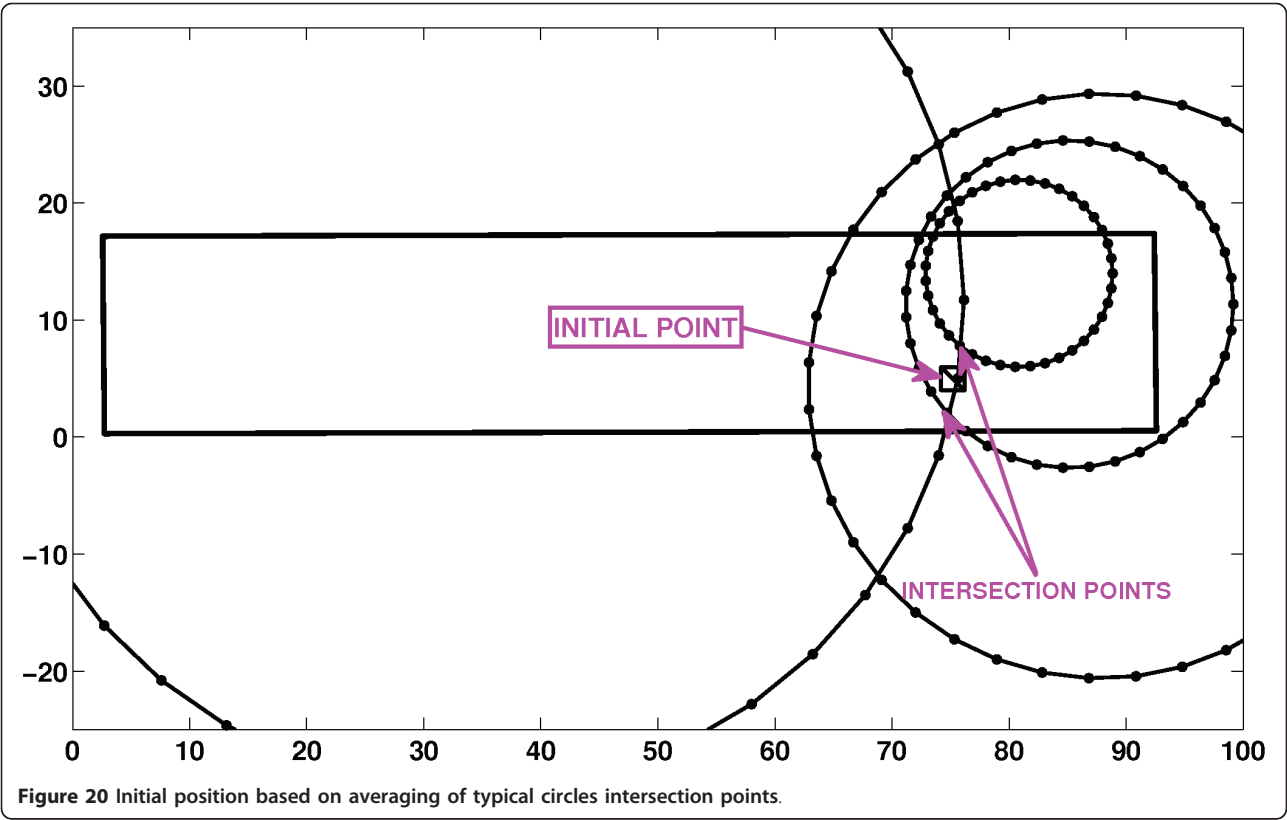
Figure 23 shows the cdf of the distance error on the preprocessing steps. The dotted line represents the error distribution for the initial step. The RSSI measurements of the test described in Section 4 at full power are converted to distances using the propagation parameters taken from the theoretical linear regression model derived from the attenuation in the IEEE802.15.4 standard when evaluated between 2.5 and 83 m (see Section 5.1). The difference between these distances and the real

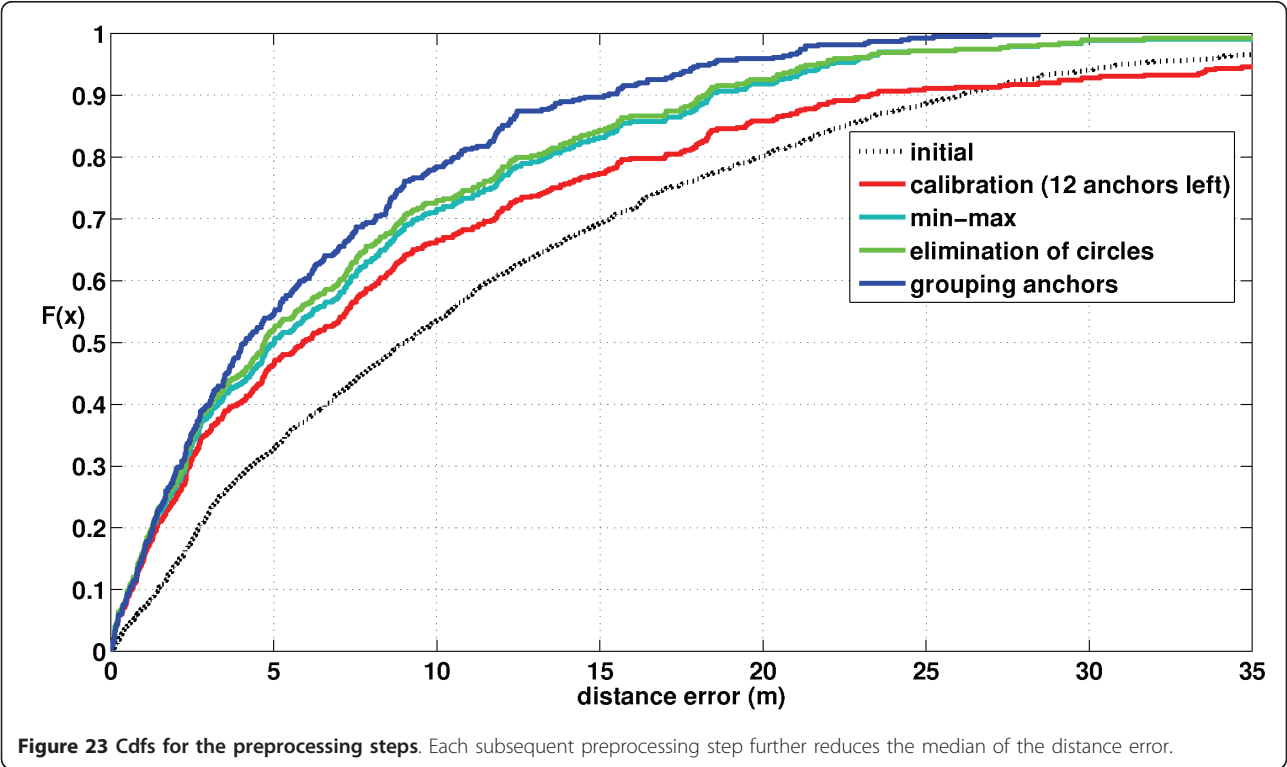
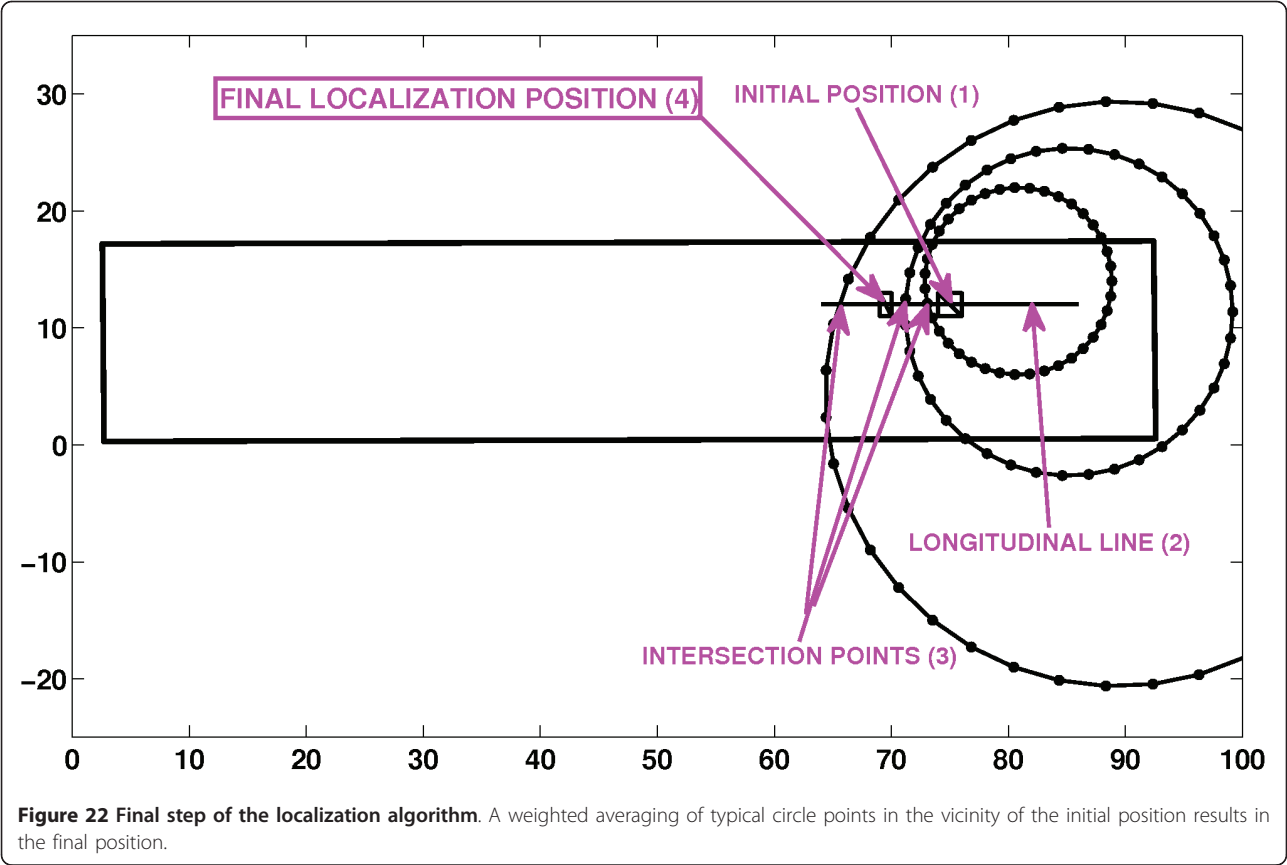
distance between the corresponding sender/receiver pair defines the (one-dimensional) distance error. The other curves present the cdf for the distance errors after subsequently applying Sections 5.2 (calibration 12 nodes left), 5.4 (min-max algorithm), 5.5 (elimination of circles), and 5.6 (grouping anchors), respectively. For the low and medium percentiles, the first preprocessing step already predicts an improvement. As can be seen at the upper right side of the figure, the calibration of anchors has a negative effect for percentiles above 90. Figure 23 also reveals that this effect is eliminated, when the too large circles are rejected. The subsequent steps of the preprocessing further improve the accuracy. It is noted that all the preprocessing steps reduced the median of the distance error from 8.95 to 4.03 m.

### 7.2. Position error

For comparison, a more conventional maximum likelihood on the position is implemented. Our building is gridded with an intergrid size of 0.50 m in each direction. Next, the exact distances  $d_{i,j}$  between the grid points  $j$  and the anchors  $i$  are calculated. Now, the anchor's RSSI measurements are used with conversion to the "calculated (tilded  $d_{i,j}$ ) distances" of Tables 2, 3, and 4. These points are compared resulting in the mean square error cost function (MMSE) (3) The most likely









location is determined by minimizing this cost function [57,58]:

$$\arg \min_{x,y} \sum_{j \in \text{anchor}(i)} (\tilde{d}_{ij} - d_{ij})^2 \quad (3)$$

In Figure 24, the error on the position is shown in our sparse anchor density (12 anchors on a surface of 1512 m<sup>2</sup> or approximately 0.008 anchors per square meter) environment. As a common reference, the end of Section 5.2 is chosen: the best anchors are selected and calibrated. The effect of the presence of the rest of the preprocessing steps (5.4 through 5.6) is studied both on our algorithm and the more conventional maximum likelihood on the position. The figure shows that there is an improvement of the complete preprocessing on our complete algorithm, and the median of the error is reduced from 6.22 to 5.29 m. Furthermore, the worst result is for MMSE without preprocessing steps 5.4-5.6. Using our min-max algorithm, eliminating circles and grouping anchors have a positive effect on the position error of the MMSE algorithm. It reduces the median from 6.66 to 6.45 m. Our complete algorithm has a lower median than the MMSE maximum likelihood algorithm with all our preprocessing steps, and it has the lowest high percentiles.

The largest errors for both the algorithms are made in the longitudinal axis of the building along the longest

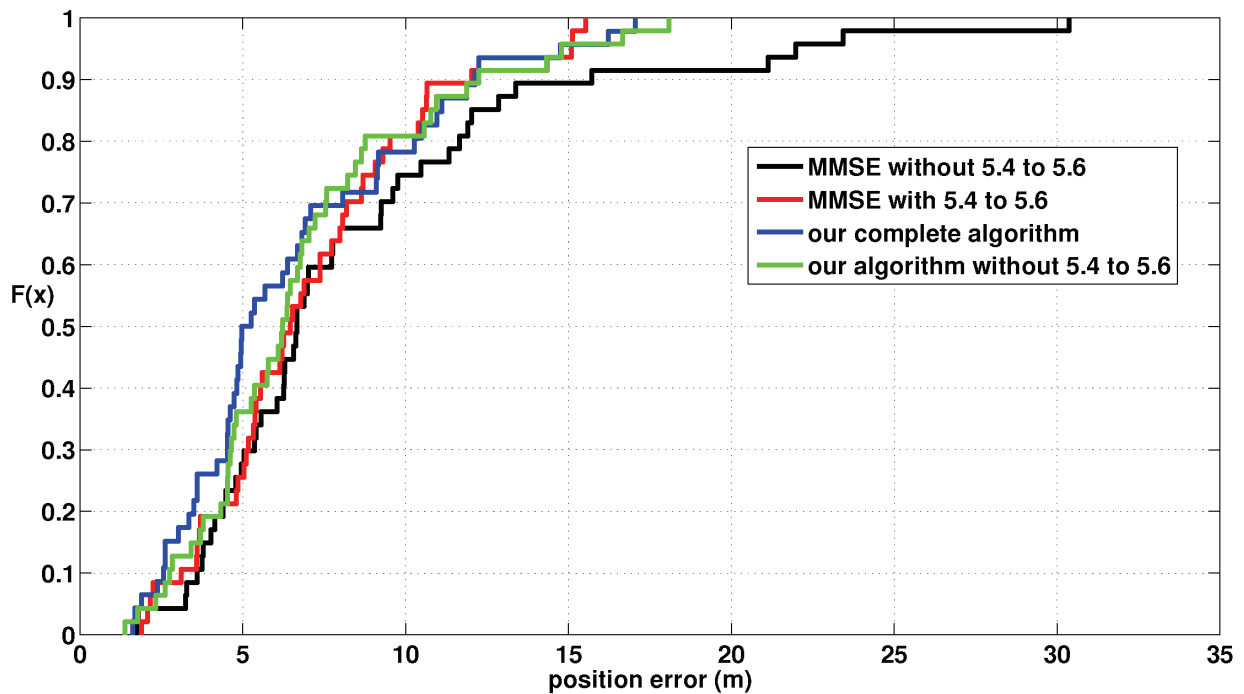
side of the rectangle. Figure 25 shows this longitudinal error for both the algorithms and illustrates the existence of many large errors for the maximum likelihood algorithm on the position. Its highest error belongs to node 2 (at a longitudinal coordinate of 91 m), situated not only in an extremity of the building (where the selected anchors are less accurate, see Figure 12) but also in the longest corridor (with the most constructive multipath fading, see Section 4).

Figure 26 shows the cdfs for the lateral error on the position. The upper graph is for our complete algorithm, and the lower graph is for the maximum likelihood on the position (without preprocessing steps 5.4-5.6). Our tests show that our algorithm tends to favor the center in a narrow rectangle, while a maximum likelihood on the position rather locates the targets on the longitudinal borders for such a geometry.

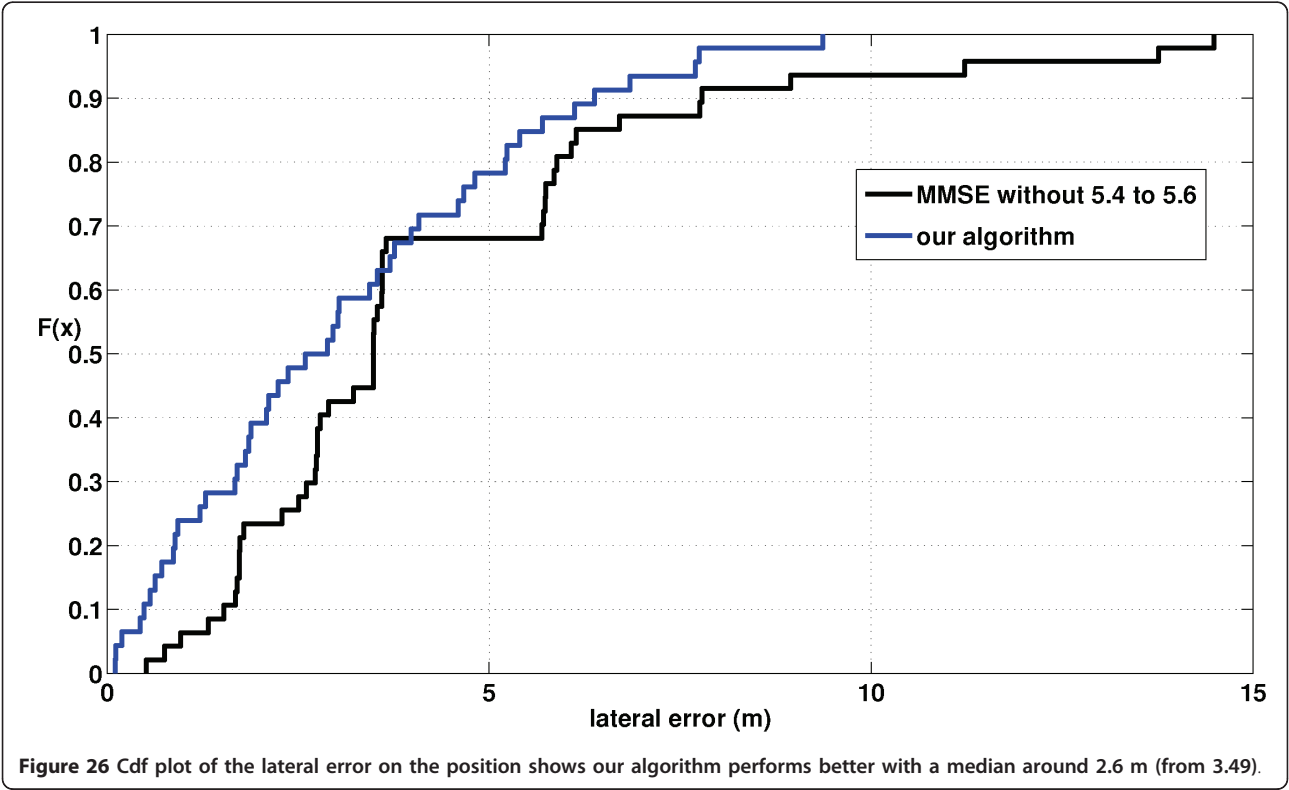
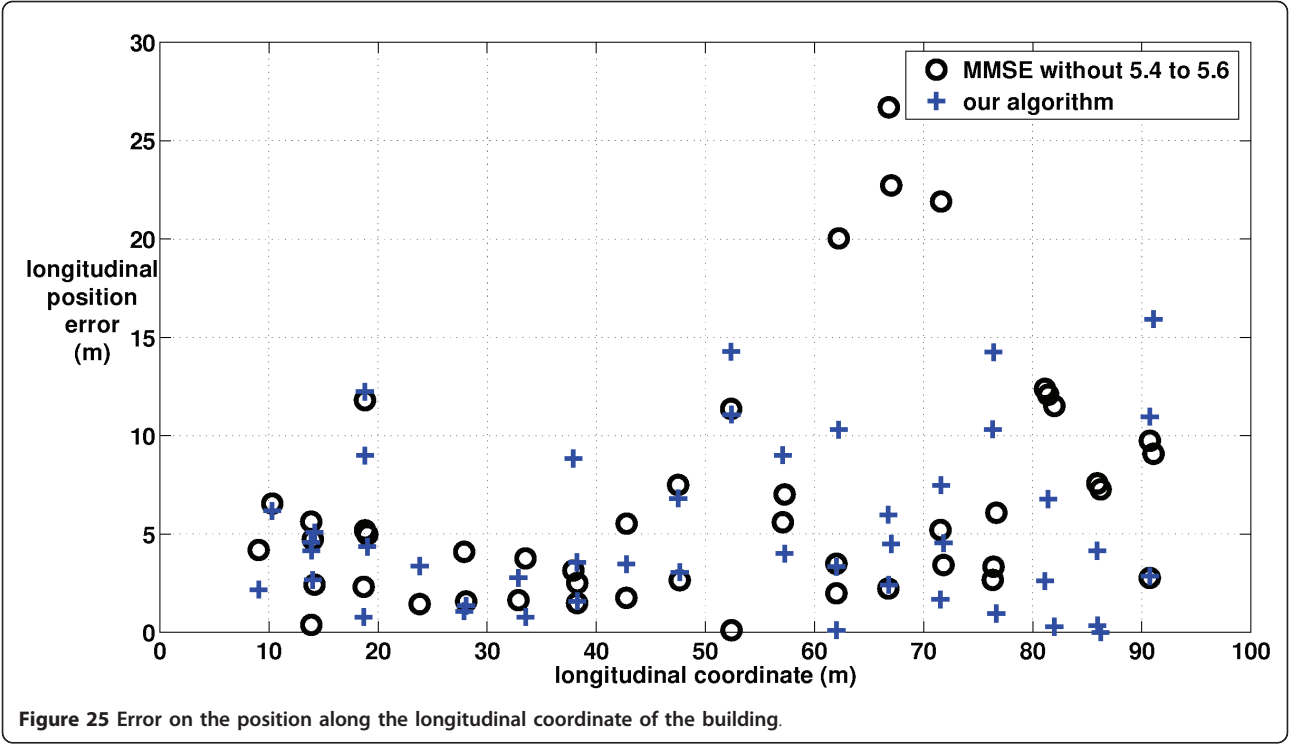
## 8. Conclusion

We have presented a new approach for localization in a realistic indoor environment, where multipath fading is highly present. Our measurements show that the preprocessing steps decrease the median of the distance error from 8.95 to 4.03 m. Furthermore, the application of our complete algorithm eliminates the outliers and obtains a median of the position error of 5.26 m.

A maximum likelihood algorithm with a mean square error cost function has a position error median of 6.66



**Figure 24** Cdfs for the error on the position. Comparison is made between our algorithm and conventional maximum likelihood methods, presented by the MMSE.



m. When our preprocessing is applied not only this median, but also the high percentiles of this algorithm are improved.

In our pre-existing sensor network with a large number of sensor nodes, the best available nodes are selected as anchors and calibrated. The whole process can be automated using standard linear regression tools. Time-consuming manual fitting and complex fingerprinting is avoided, thus, making it possible to do real-time localization in future dynamic wireless indoor environments.

#### Abbreviations

EE: environment emulator; IBBT: Interdisciplinary Institute for Broadband Technology; LQI: link-quality indication; RSSI: received signal strength indication; WSN: wireless sensor network.

#### Acknowledgements

The authors wish to thank the IBBT and all the designers of the W-ilab.t test bed in general and, Bart Jooris in particular for his support with the experiments.

#### Author details

<sup>1</sup>Faculty of Applied Engineering Sciences, University College Ghent, Ghent, Belgium <sup>2</sup>Department of Information Technology (INTEC), IBBT, Ghent University, Ghent, Belgium

#### Competing interests

The authors declare that they have no competing interests.

Received: 1 December 2010 Accepted: 12 July 2011

Published: 12 July 2011

#### References

1. P Barsocchi, S Lenzi, S Chessa, G Giunta, A novel approach to indoor RSSI localization by automatic calibration of the wireless propagation model, in *Proceedings of the IEEE Vehicular Technology Conference (VTC)*, Barcelona, Spain (April 2009)
2. K Lorincz, M Welsh, MoteTrack: a robust, decentralized approach to RF-based location tracking, in *Proceedings of the International Workshop on Location and Context-Awareness (LoCA 2005)*, Pervasive 2005 (May 2005)
3. MK Awad, KT Wong, L Zheng-bin, An integrated overview of the open literature's empirical data on the indoor radiowave channel's delay properties. *IEEE Trans Antennas Propag.* **56**(5), 1451–1468 (2008)
4. D Lymberopoulos, Q Lindsey, A Savvides, An empirical characterization of radio signal strength variability in 3-D IEEE 802.15.4 networks using monopole antennas, in *Proceedings of the Third European Workshop on Wireless Sensor Networks*, Zurich, Switzerland (February 2006)
5. A Azenha, L Peneda, A Carvalho, Error analysis in indoors localization using ZigBee wireless networks. in *Proceedings of the 36th Annual Conference of the IEEE Industrial Electronics Society (IECON 2010)*, Phoenix, AZ (7–10 November 2010), pp. 2193–2197
6. S Alikhani, University of Ottawa website. Available at <http://www.site.uottawa.ca/~casteig/files/csi5140-shafagh-alikhani.ppt>
7. R Peng, ML Sichert, Angle of arrival localization for wireless sensor networks, in *Proceedings of the Third Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks*, Reston, VA (September 2006)
8. S Venkatraman, J Caffery, H You, A novel ToA location algorithm using LoS range estimation for NLoS environments. *IEEE Trans Veh Technol.* **53**(5), 1515–1524 (2004). doi:10.1109/TVT.2004.832384
9. A Catovic, Z Sahinoglu, Hybrid TOA/RSS and TDOA/RSS location estimation schemes for short-range wireless networks. *Bechtel Telecommun Tech J (BTTJ)*. **2**(2), 77–84 (2004)
10. EO Dijk, Indoor ultrasonic position estimation using a single base station. PhD thesis, Technische Universiteit Eindhoven (2004)
11. J Hallberg, M Nilsson, Positioning with bluetooth, IrDA and RFID. Master's thesis, Lulea University of Technology, Sweden, March 2002
12. BF Rolfe, SW Ekanayake, PN Pathirana, M Palaniswami, Localization with orientation using RSSI measurements: RF map based approach, in *Proceedings of the Third International Conference on Intelligent Sensors, Sensor Networks and Information (ISSNIP)*, (December 2007), pp. 311–316
13. C Tseng, S Cheng, Location management scheme with WLAN positioning algorithm for integrated wireless networks. *Comput Commun (COMCOM)*. **31**, 4304–4311 (2008). doi:10.1016/j.comcom.2008.07.007
14. A Kushki, KN Plataniotis, AN Venetsanopoulos, Indoor positioning with wireless local area networks (WLAN). *Encyclopedia of GIS*. 566–571 (2008)
15. Z Sahinoglu, S Gezici, I Güvenc, *Ultra-Wideband Positioning Systems Theoretical Limits, Ranging Algorithms, and Protocols*, (DoCoMo Communications Laboratories USA, CA, 2008) ISBN-13: 9780521873093
16. K Yu, I Oppermann, UWB positioning for wireless embedded networks, in *Proceedings of the IEEE Radio and Wireless Conference (RAWCON)*, Atlanta, GA (19–22 September 2004). pp. 459–462
17. N Alsindi, K Pahlavan, Cooperative localization bounds for indoor ultra-wideband wireless sensor networks. *EURASIP J Adv Signal Process.* 1–13 (2008)
18. NB Priyantha, Miu AKL, H Balakrishnan, S Teller, The cricket compass for context-aware mobile applications, in *Proceedings of the 6th ACM MOBICom Conference*, Rome, Italy (July 2001)
19. LE Miller, PF Wilson, NP Bryner, MH Francis, JR Guerrieri, DW Stroup, L Klein-Berndt, RFID-assisted indoor localization and communication for first responders, in *Proceedings 2006 International Symposium on Advanced Radio Tech (ISART)*, Boulder, CO (March 2006), pp. 83–91
20. IEEE Standard for Information technology–Telecommunications and information exchange between systems–Local and metropolitan area networks–Specific requirements–Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications, Std IEEE802.11-2007 (2007)
21. C Connell, ScenSor: an IEEE802.15.4a UWB compliant chip ripe for energy harvesting. in *Energy Harvesting & Storage Europe 2009*, Cambridge, UK <http://www.idtechex.com/events/presentations/scensor-an-ieee802-15-4a-uwb-compliant-chip-ripe-for-energy-harvesting-tapes-out-q4-09-001925.asp> (4 June 2009)
22. Y Shen, MZ Win, Fundamental limits of wideband localization–Part I: a general framework. *IEEE Trans Inf Theory*. **56**(10), 4956–4980 (2010)
23. Y Shen, H Wymeersch, MZ Win, Fundamental limits of wideband localization–Part II: cooperative networks. *IEEE Trans Inf Theory*. **56**(10), 4981–5000 (2010)
24. Y Shen, MZ Win, On the accuracy of localization systems using wideband antenna arrays. *IEEE Trans Commun.* **58**(1), 270–280 (2010)
25. D Dardari, A Conti, U Fener, A Giorgetti, MZ Win, Ranging with ultrawidebandwidth signals in multipath environments. *Proc IEEE*. **97**(2), 404–426 (2009)
26. S Marano, WM Gifford, H Wymeersch, MZ Win, NLOS identification and mitigation for localization based on UWB experimental data. *IEEE J Sel Areas Commun.* **28**(7), 1026–1035 (2010)
27. H Liu, H Darabi, P Banerjee, J Liu, Survey of wireless indoor positioning techniques and systems. *IEEE Trans Syst Man Cyber C (TSMC)*. **37**(6), 1067–1080 (2007)
28. S Tennina, Renzo M Di, F Graziosi, F Santucci, Locating ZigBee nodes using the TI's CC2431 location engine: a testbed platform and new solutions for positioning estimation of WSNs in dynamic indoor environments, in *Proceedings of the First ACM International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments, MELT 2008*, San Francisco, CA (September 2008), pp. 37–42
29. H Cho, M Kang, JH Park, B Park, H Kim, Performance analysis of location estimation algorithm in ZigBee networks using received signal strength. in *Proceedings of the 21st IEEE International Conference on Advanced Information Networking and Applications Workshops (AINA-07)*, vol. 2. Niagara Falls, Canada (May 2007), pp. 302–306
30. S Capkun, M Hamdi, J-P Hubaux, GPS-free positioning in mobile ad-hoc networks, in *Proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS-34)*, Wailea Maui, HI (January 2001)
31. D Goldenberg, A Krishnamurthy, WC Maness, YR Yang, A Young, AS Morse, A Savvides, BDO Anderson, Network localization in partially localizable networks, in *Proceedings of IEEE Infocom 2005*, Miami, FL (March 2005)
32. T He, C Huang, BM Blum, JA Stankovic, TF Abdelzaher, Range-free localization schemes in large scale sensor networks, in *Proceedings of the*

- ACM International Conference on Mobile Computing and Networking (MobiCom 2003) (2003)
33. H Wymeersch, J Lien, MZ Win, Cooperative localization in wireless networks. *Proc IEEE*. **97**(2), 427–450 (2009)
  34. G Chandrasekaran, MA Ergin, J Yang, S Liu, Y Chen, M Gruteser, RP Martin, Empirical evaluation of the limits on localization using signal strength, in *Proceedings of the 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON'09)*, Rome, Italy (June 2009), pp. 333–341
  35. D Madigan, E Elnahrawy, RP Martin, Ju W-H, P Krishnan, AS Krishnakumar, Bayesian indoor positioning systems, in *Proceedings of IEEE Infocom*, Miami, FL (March 2005)
  36. M Nicoli, C Morelli, V Rampa, U Spagnolini, HMM-based tracking of moving terminals in dense multipath indoor environments, in *Proceedings of the EURASIP European Signal Processing Conference (EUSIPCO'05)*, Antalya, Turkey, (September 2005)
  37. T Roos, P Myllymaki, H Tiiri, A statistical modeling approach to location estimation. *IEEE Trans Mobile Comput.* **1**(1), 59–69 (2002). doi:10.1109/TMC.2002.1011059
  38. M Saxena, P Gupta, BN Jain, Experimental analysis of RSSI-based location estimation in wireless sensor networks, in *Proceedings of the 3rd International Conference on Communication Systems Software and Middleware and Workshops (COMSWARE)*, Bangalore, India (5–10 January 2008) pp. 503–510
  39. P Barsocchi, S Lenzi, S Chessa, G Giunta, Virtual calibration for RSSI-based indoor localization with IEEE 802.15.4. in *Proceedings of the IEEE International Conference on Communications (ICC)*, Dresden, Germany (14–18 June 2009) pp. 5–10
  40. L Tytgat, B Jooris, P De Mil, B Latré, I Moerman, P Demeester, Demo abstract: WiLab, a real-life wireless sensor testbed with environment emulation, in *Proceedings of the 6th European Conference on Wireless Sensor Networks (EWSN)*, Cork, Ireland (11–13 February 2009)
  41. F Vanheel, J Verhaevert, I Moerman, Study on distance of interference sources on wireless sensor network, in *Proceedings of the European Microwave Conference (EuMC)*, Amsterdam, The Netherlands (27–31 October 2008), pp. EuMC10-1.1–4
  42. EB Hamida, G Chelius, Investigating the impact of human activity on the performance of wireless networks—an experimental approach, in *Proceedings of the 2010 IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks (WoWMoM)*, Montreal, Canada (14–17 June 2010), pp. 1–8
  43. N Yarkoni, N Blaunstein, Prediction of propagation characteristics in indoor radio communication environments. *Prog Electromag Res PIER*. **59**, 151–174 (2006)
  44. K Fujimoto, JR James, *Mobile Antenna Systems Handbook*, 2nd edn. (Artech House, 2001), ISBN-13: 9781580530071
  45. Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs), IEEE Std 802.15.4-2006 (2006)
  46. JS Bendat, AG Piersol, *Random Data, Analysis and Measurement Procedures*, 2nd edn. (John Wiley & Sons, New York, 1986), ISBN-10: 0471317330
  47. P Pivato, L Fontana, L Palopoli, D Petri, Experimental assessment of a RSSI-based localization algorithm in indoor environment, in *Instrumentation and Measurement Technology Conference (I2MTC)*, 2010 IEEE (3–6 May 2010), pp. 416–421
  48. H Hashemi, The indoor radio propagation channel, in *Proceedings of the IEEE Vehicular Technology Conference (VTC 92)*, Denver, CO (May 1992)
  49. P Barsocchi, S Lenzi, S Chessa, F Furfari, Automatic virtual calibration of range-based indoor localization systems. *Wirel Commun Mobile Comput* [online]. Available at <http://onlinelibrary.wiley.com/doi/10.1002/wcm.1085/pdf>
  50. N Ha, K Han, Positioning method for outdoor systems in wireless sensor networks, in *Proceedings of the 21st International Symposium on Computer and Information Sciences (ISCIS 2006)*, Istanbul, Turkey (November 2006), pp. 783–792
  51. L Doherty, K Pister, L El Ghaoui, Convex position estimation in wireless sensor networks, in *Proceedings of IEEE Infocom 2001*, Anchorage, AK (April 2001), pp. 1655–1663
  52. J Neter, W Wasserman, M Kutner, *Applied Linear Statistical Models*, 5th edn. (MacGraw Hill, New York, 2005), ISBN-10: 0072386886
  53. S Mazuelas, A Bahillo, RM Lorenzo, P Fernandez, FA Lago, E Garcia, J Blas, EJ Abril, Robust indoor positioning provided by real-time rssi values in unmodified wlan networks. *IEEE J Sel Topics Signal Process.* **3**(5), 821–831 (2009)
  54. K Whitehouse, D Culler, Calibration as parameter estimation in sensor networks, in *Proceedings of ACM International Workshop on Wireless Sensor Networks and Applications (WSNA'02)*, Atlanta, GA (28 September 2002), pp. 59–67
  55. Y Kong, Y Kwon, G Park, Practical robust localization over obstructed interferences in wireless sensor networks, in *2009 Digest of Technical Papers International Conference on Consumer Electronics*, Las Vegas, NV (January 2009), pp. 1–2
  56. M Terwilliger, CR Coullard, AK Gupta, Localization in ad hoc and sensor wireless networks with bounded errors, in *Proceedings of the 15th International Conference on High Performance Computing (HiPC)*, Bangalore, India (2008), pp. 295–308
  57. X An, J Wang, VR Prasad, IGMM Niemegeers, OPT: online person tracking system for context-awareness in wireless personal network, in *Proceedings of the 2nd international workshop on Multi-hop ad hoc networks: from theory to reality (2006)*, Florence, Italy (May 2006), pp. 47–54
  58. N Patwari, AO Hero, M Perkins, NS Correal, RJ O'Dea, Relative location estimation in wireless sensor networks. *IEEE Trans Signal Process.* **51**, 2137–2148 (2003). doi:10.1109/TSP.2003.814469

doi:10.1186/1687-1499-2011-38

**Cite this article as:** Vanheel et al.: Automated linear regression tools improve RSSI WSN localization in multipath indoor environment. *EURASIP Journal on Wireless Communications and Networking* 2011 **2011**:38.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)